GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 3.63509 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-21
Perfect score: 1093
Sequence: 1 MLLCDIGNSNANFLDDNKYF.....KLLIFRGMKKIIKENPNLLY 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	96	8.8	356	1	LPXB_AQUAE	O67420 aquifex aeo
2	94	8.6	321	1	VG16_BPB03	Q37897 bacteriopha
3	92.5	8.5	2652	1	RRPB_IBVB	P26314 avian infec
4	90	8.2	1272	1	UBP2_YEAST	Q01476 saccharomyc
5	89	8.1	622	1	MUTL_CLOAB	Q97120 clostridium
6	89	8.1	1024	1	Y075_MYCGE	P47321 mycoplasma
7	88.5	8.1	901	1	GREA_BORBU	O51157 borrelia bu
8	87	8.0	319	1	COAA_RHILO	Q98cs9 rhizobium l
9	86.5	7.9	448	1	MQO_CAMJE	P56954 campylobact
10	85.5	7.8	1500	1	SSP5_STRGN	P16952 streptococc
11	84.5	7.7	442	1	TRME_MYCGE	P47254 mycoplasma
12	83	7.6	269	1	FLIR_BORBU	Q44907 borrelia bu
13	82.5	7.5	279	1	PPNK_BORBU	O51291 borrelia bu
14	82.5	7.5	708	1	PTA_BUCAI	P57273 buchnera ap
15	82	7.5	249	1	YBFT_BACSU	O31458 bacillus su
16	82	7.5	613	1	CGAA_CLOBI	Q45882 clostridium
17	82	7.5	1513	1	STU1_YEAST	P38198 saccharomyc
18	82	7.5	2278	1	FAB1_YEAST	P34756 saccharomyc
19	81.5	7.5	2515	1	TUD_DROME	P25823 drosophila
20	81	7.4	204	1	TRPF_SULSO	P50386 sulfolobus
21	81	7.4	318	1	RCN2_RAT	Q62703 rattus norv
22	81	7.4	621	1	HTPG_HELPY	P56116 helicobacte
23	80.5	7.4	292	1	RNH3_LACLA	Q9cdg3 lactococcus
24	80.5	7.4	1024	1	RPOB_PLAFA	P21421 plasmodium
25	80	7.3	808	1	SECA_MYCPN	P75559 mycoplasma
26	79.5	7.3	476	1	RF3_SACBA	P05512 saccharomyc
27	79.5	7.3	621	1	HTPG_HELPJ	Q9zmm2 helicobacte
28	79.5	7.3	867	1	PROM_MOUSE	O54990 mus musculu
29	79.5	7.3	1162	1	BXEN_CLOBO	P46082 clostridium
30	79.5	7.3	1162	1	BXEN_CLOBU	Q06366 clostridium
31	79	7.2	314	1	MTH1_HAEPA	P29538 haemophilus
32	79	7.2	667	1	NAH9_CAEEL	P35449 caenorhabdi
33	79	7.2	964	1	DPOL_CBEPV	P30319 choristoneu
34	78.5	7.2	389	1	ACDP_MYCLE	P46703 mycobacteri
35	78.5	7.2	481	1	TLDD_ECOLI	P46473 escherichia
36	78.5	7.2	1034	1	ENTK_PIG	P98074 sus scrofa
37	78.5	7.2	1630	1	MSP1_PLAFK	P04932 plasmodium
38	78.5	7.2	1639	1	MSP1_PLAFW	P04933 plasmodium
39	78	7.1	392	1	UDG5_ECOLI	Q47329 escherichia
40	78	7.1	396	1	RMAR_YEAST	P02381 saccharomyc
41	78	7.1	587	1	UVSW_BPT4	P20703 bacteriopha
42	78	7.1	1025	1	PMPF_CHLMU	Q9p146 chlamydia m
43	78	7.1	1161	1	RPO2_FOWPV	Q9j544 fowlpox vir
44	78	7.1	2230	1	POLG_HPAVS	P14553 simian hepa
45	77.5	7.1	187	1	Y163_METJA	Q57627 methanococc

ALIGNMENTS

RESULT 1
LPXB_AQUAE
ID LPXB_AQUAE STANDARD; PRT; 356 AA.
AC O67420;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipid-A-disaccharide synthase (EC 2.4.1.182).
GN LPXB OR AQ_1427.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: CONDENSATION OF UDP-2,3-DIACYLGLUCOSAMINE AND
CC 2,3-DIACYLGLUCOSAMINE-1-PHOSPHATE TO FORM LIPID A DISACCHARIDE,
CC A PRECURSOR OF LIPID A, A PHOSPHORYLATED GLYCOLIPID THAT ANCHORS
CC THE LIPOPOLYSACCHARIDE TO THE OUTER MEMBRANE OF THE CELL
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-2,3-bis(3-hydroxytetradecanoyl)glucosamine
CC + 2,3-bis(3-hydroxytetradecanoyl)-beta-D-glucosaminyl 1-phosphate
CC = UDP + 2,3-bis(3-hydroxytetradecanoyl)-D-glucosaminyl-1,6-beta-D-
CC 2,3-bis(3-hydroxytetradecanoyl)-beta-D-glucosaminyl 1-phosphate.
CC -!- PATHWAY: Lipid A biosynthesis.
CC -!- SIMILARITY: BELONGS TO THE LPXB FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE000740; AAC07386.1; -.
DR InterPro; IPR003835; LpxB.
DR Pfam; PF02684; LpxB; 1.
DR TIGRFAMs; TIGR00215; lpxB; 1.
KW Transferase; Glycosyltransferase; Lipid A biosynthesis;
KW Lipid synthesis; Complete proteome.
SQ SEQUENCE 356 AA; 41300 MW; 1B4CFBA5F409CD68 CRC64;

Query Match 8.8%; Score 96; DB 1; Length 356;
Best Local Similarity 26.1%; Pred. No. 0.74;
Matches 48; Conservative 29; Mismatches 63; Indels 44; Gaps 9;

Qy 34 QKIFYI----NVNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDRIAACYTIEDGVVVD 89

```
              :|||      : : :: | ||  : |   | | | |   : :|:  :|    |    :::
Db          2 KKIFLSLADRSASNYVYEILK--EGFEEYEIYGLTDEKLEKIGVKSVARYSEISTVGLIE 59

Qy         90 AGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTM---- 145
               |                   || :  : |||  |   ||:  :| :: || |   :
Db         60 A------------------LPKVFKFLKIYRKILKNLKNT-DTLIACDA-PALNLRLIK 98

Qy        146 DALSYGVFKGIYLLIKD--AAQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKE 203
              ||   || : || :     | : |:       : :||| || |    :|    ||| ::
Db         99 DARKLGVKRIIYFISPQVWAWKPKR-------AEIIANYCDHVI---VILPFEKKIYRK 147

Qy        204 NPNL 207
               |||
Db        148 FPNL 151


RESULT 2
VG16_BPB03
ID   VG16_BPB03     STANDARD;      PRT;   321 AA.
AC   Q37897;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Encapsidation protein (Late protein GP16).
GN   16.
OS   Bacteriophage B103.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
OC   phi-29-like viruses.
OX   NCBI_TaxID=10778;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98019084; PubMed=9358052;
RA   Pecenkova T., Benes V., Paces J., Vlcek C., Paces V.;
RT   "Bacteriophage B103: complete DNA sequence of its genome and
RT   relationship to other Bacillus phages.";
RL   Gene 199:157-163(1997).
CC   -!- FUNCTION: REQUIRED FOR THE GENOME ENCAPSIDATION REACTION (BY
CC       SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X99260; CAA67647.1; -.
KW   Late protein; ATP-binding.
FT   NP_BIND      13     20       ATP (POTENTIAL).
SQ   SEQUENCE   321 AA;  37651 MW;  6BB405E6E07BF1F3 CRC64;

  Query Match           8.6%;  Score 94;  DB 1;  Length 321;
  Best Local Similarity  24.1%;  Pred. No. 0.96;
  Matches   38;  Conservative   24;  Mismatches   40;  Indels   56;  Gaps    9;

Qy         14 LDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGI 73
              :|  :|  :::|   || |:  :|       :::  || |        |:|  :|:|: :
Db        198 IDGTEYGEMSLDN--EFVNDSDVF-------IEKRSKNSK--------FVFSIVYKGMRM 240

Qy         74 DRIAACYTIEDGVVVDAGS----------------AITIDIISNSIHLGGFILPGIANY 116
                 || ||                          |:| | :       | ||  ::||
Db        241 -----------GVWVDTQQMLLYLTTDHDPSTKNVYALTADDLEE-----GMIL--VSNY 282

Qy        117 KKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFK 154
              || | ||     :  | |:  |    : |  : | :||
Db        283 KKNY-HIRKLCSAFMNGQLRFD---NQLMRTIGYEMFK 316


RESULT 3
RRPB_IBVB
ID   RRPB_IBVB      STANDARD;      PRT;  2652 AA.
AC   P26314;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   RNA-directed RNA polymerase (ORF1B) (EC 2.7.7.48).
GN   F2.
OS   Avian infectious bronchitis virus (strain Beaudette) (IBV).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC   Coronaviridae; Coronavirus.
OX   NCBI_TaxID=11122;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87111468; PubMed=3027249;
RA   Boursnell M.E.G., Brown T.D.K., Foulds I.J., Green P.F., Tomley F.M.,
RA   Binns M.M.;
RT   "Completion of the sequence of the genome of the coronavirus avian
RT   infectious bronchitis virus.";
RL   J. Gen. Virol. 68:57-77(1987).
CC   -!- FUNCTION: THE RNA DEPENDENT RNA POLYMERASE OF CORONAVIRUSES IS
CC       A MULTIFUNCTIONAL PROTEIN: IT CONTAINS THE ACTIVITIES NECESSARY
CC       FOR THE TRANSCRIPTION OF NEGATIVE STRANDED RNA, LEADER RNA,
CC       SUBGENOMIC MRNAS AND PROGENY VIRION RNA.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- MISCELLANEOUS: THIS PROTEIN IS EXPRESSED BY AN EFFICIENT RIBOSOMAL
CC       FRAMESHIFTING MECHANISM. RIBOSOMAL FRAMESHIFTING IS AN ELEGANT
CC       MECHANISM FOR REGULATING THE SYNTHESIS OF SEVERAL PROTEINS IN A
CC       WELL BALANCED MANNER.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94356; AAA46224.1; -.
DR   EMBL; M95169; AAA70234.1; -.
DR   PIR; B33094; VFIHB2.
DR   InterPro; IPR000606; Viral_helicase1.
DR   Pfam; PF01443; Viral_helicase1; 1.
KW   Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT   NP_BIND    1173   1180       ATP (BY SIMILARITY).
SQ   SEQUENCE   2652 AA;  300617 MW;  F5D7DBFD09D1E29D CRC64;

  Query Match           8.5%;  Score 92.5;  DB 1;  Length 2652;
  Best Local Similarity  24.0%;  Pred. No. 14;
  Matches   54;  Conservative   39;  Mismatches   83;  Indels   49;  Gaps   12;

Qy          3 LCDIGNSNANFLDDN----KYFTLNIDQF-LEFKNEQKIFYINVNEHLKEHLK-NQKNFI 56
              ||   | | :   ||    :| | |:  | |   |   ::   ||:|     | :: :|
Db       1867 LCMFWNCNVDCYPDNSLVCRYDTRNLSVFNLPGCNGGSLY---VNKHAFYTPKFDRISFR 1923

Qy         57 NLE--PYFLFDTIYQGLGIDRIAACYTIE-DGVVVDAGSAITIDIISNSIHLGGFILPGI 113
              ||:  |:| :|:          : | ||: |||  |  |  | | |:   ::|| :
Db       1924 NLKAMPFFFYDS----------SPCETIQVDGVAQDLVSLATKDCIT-KCNIGGAVCKKH 1972

Qy        114 AN-YKKIYSHISPRLKSEFNTQVSLDAFPQ---------KTMDALSYGVFKGIYLLIKDA 163
              |  | :  :  :  : : |   |:    |          :::| ::| ::|| :    ||
Db       1973 AQMYAEFVTSYNAAVTAGFTFWVTNKLNPYNLWKSFSALQSIDNIAYNMYKGGHY---DA 2029

Qy        164 AQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKENPNLL 208
                :      ||  |             :| :|::| :  |   |
Db       2030 IAGEMPTVITGDKVF------------VIDQGVEKAVFVNQTTL 2061


RESULT 4
UBP2_YEAST
ID   UBP2_YEAST     STANDARD;      PRT;  1272 AA.
AC   Q01476; Q99357;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
```

```
DE   Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
DE   thiolesterase 2) (Ubiquitin-specific processing protease 2)
DE   (Deubiquitinating enzyme 2).
GN   UBP2 OR YOR124C OR O3281 OR YOR3281C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93054674; PubMed=1429680;
RA   Baker R.T., Tobias J.W., Varshavsky A.;
RT   "Ubiquitin-specific proteases of Saccharomyces cerevisiae. Cloning of
RT   UBP2 and UBP3, and functional analysis of the UBP gene family.";
RL   J. Biol. Chem. 267:23364-23375(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=97060020; PubMed=8904341;
RA   Wiemann S., Rechmann S., Benes V., Voss H., Schwager C., Vlcek C.,
RA   Stegemann J., Zimmermann J., Erfle H., Paces V., Ansorge W.;
RT   "Sequencing and analysis of 51 kb on the right arm of chromosome XV
RT   from Saccharomyces cerevisiae reveals 30 open reading frames.";
RL   Yeast 12:281-288(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97344368; PubMed=9200815;
RA   Voss H., Benes V., Andrade M.A., Valencia A., Rechmann S., Teodoru C.,
RA   Schwager C., Paces V., Sander C., Ansorge W.;
RT   "DNA sequencing and analysis of 130 kb from yeast chromosome XV.";
RL   Yeast 13:655-672(1997).
CC   -!- FUNCTION: HAS AN ATP-INDEPENDENT ISOPEPTIDASE ACTIVITY, CLEAVING
CC       AT THE CARBOXYL TERMINUS OF THE UBIQUITIN MOIETY IN NATURAL OR
CC       ENGINEERED LINEAR FUSION PROTEINS, IRRESPECTIVE OF THEIR SIZE OR
CC       THE PRESENCE OF AN AMINO-TERMINAL EXTENSION TO UBIQUITIN.
CC   -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC       ubiquitin + a thiol.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M94916; AAA35190.1; -.
DR   EMBL; X94335; CAA64043.1; -.
DR   EMBL; X90518; CAA62120.1; -.
DR   EMBL; Z75032; CAA99323.1; -.
DR   PIR; S27466; S27466.
DR   PIR; A44450; A44450.
DR   MEROPS; C19.003; -.
DR   SGD; S0005650; UBP2.
DR   InterPro; IPR001394; UCH-2.
DR   Pfam; PF00442; UCH-1; 1.
DR   Pfam; PF00443; UCH-2; 1.
DR   PROSITE; PS00972; UCH_2_1; 1.
DR   PROSITE; PS00973; UCH_2_2; 1.
DR   PROSITE; PS50235; UCH_2_3; 1.
KW   Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT   ACT_SITE    745    745       BY SIMILARITY.
FT   ACT_SITE   1201   1201       BY SIMILARITY.
FT   ACT_SITE   1209   1209       BY SIMILARITY.
FT   CONFLICT    658    665       MISSING (IN REF. 1).
SQ   SEQUENCE   1272 AA;  146354 MW;  6D106539AE5C5F3F CRC64;

  Query Match             8.2%;  Score 90;  DB 1;  Length 1272;
  Best Local Similarity   22.5%;  Pred. No. 9.8;
  Matches   54;  Conservative   24;  Mismatches   78;  Indels   84;  Gaps   13;

Qy          4 CDIGNSN-ANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKN-------- 54
              |:|   | |  |   || | |  |                    || | ||
Db        525 CNISELNEATLLSIYKYETSNKSQV----------------TSNHLTNLKNALRLLAKY 567

Qy         55 --------FINLEPYFLFDTIYQGLGIDR------IAACYTIE----DGVVVDAGSAI-T 95
                    ::: |||      |  | ||       |   |:::     |: :|   |: |
Db        568 TKSDKLKFYVDHEPYRALSQAYDTLSIDESVDEDIIKTAYSVKINDSPGLKLDCDRALYT 627

Qy         96 IDIISNSIHLGGFI---LPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGV 152
              | |   |: |  |:    |  :||       |  | ::   : |    :  |    :
Db        628 IAISKRSLDLFNFLTEECPQFSNY------YGPE-KLDYQEALKLLQVNENASDETILKI 680

Qy        153 FKGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIY--DKLLIFR-GMKKI-IKENPNLL 208
              ||                          :||  :|  |: || |  : || |: |  |:
Db        681 FK-----------------------QKWFDENVYEPDQFLILRAALTKISIERNSTLI 715


RESULT 5
MUTL_CLOAB
ID   MUTL_CLOAB     STANDARD;      PRT;   622 AA.
AC   Q97I20;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA mismatch repair protein mutL.
GN   MUTL OR CAC1836.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
CC   -!- FUNCTION: This protein is involved in the repair of mismatches in
CC       DNA. It is required for dam-dependent methyl-directed DNA mismatch
CC       repair. May act as a "molecular matchmaker", a protein that
CC       promotes the formation of a stable complex between two or more
CC       DNA-binding proteins in an ATP-dependent manner without itself
CC       being part of a final effector complex (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE007692; AAK79800.1; -.
DR   InterPro; IPR003594; ATPbind_ATPase.
DR   InterPro; IPR002099; DNA_mis_repair.
DR   InterPro; IPR004359; HIS_KIN_sig.
DR   Pfam; PF01119; DNA_mis_repair; 1.
DR   Pfam; PF02518; HATPase_c; 1.
DR   SMART; SM00387; HATPase_c; 1.
DR   TIGRFAMs; TIGR00585; mutl; 1.
DR   PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW   DNA repair; Complete proteome.
SQ   SEQUENCE   622 AA;  71064 MW;  D372ED449865011A CRC64;

  Query Match             8.1%;  Score 89;  DB 1;  Length 622;
  Best Local Similarity   22.5%;  Pred. No. 5.2;
  Matches   56;  Conservative   39;  Mismatches   72;  Indels   82;  Gaps   14;

Qy          1 MLLCDIGNSNANFLDDNK--YFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINL 58
```

```
Db        262 LITAAVENAVKSFMMINKFPPFIIFLDIFPEF--------VDVNVH---PTKSEVKFQNE 310

Qy         59 EPYF--LFDTIYQGLGID-----RIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFI-- 109

Db        311 RDIFKIIFDTVHEGIRNSLKESFKVEALKEEEDKLFDIKEDVITKNEIKHDDKIDGYIKK 370

Qy        110 --------LP-----GIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGI 156

Db        371 DSFPVPVQIPIDLKRPIENYDN---------STKENKDRSFDDFREKD------------ 409

Qy        157 YLLIKDAAQNKKLY--FTGGD---------GQF-----LANYFD-------HAIYDKLLI 193

Db        410 --IIKETSQEKETYDIITNKKAKFPELRVIGQFNNTYILAESFEELYIIDQHAAHEKIL- 466

Qy        194 FRGMKKIIK 202

Db        467 FEKYREDIK 475
```

RESULT 6
Y075_MYCGE
ID Y075_MYCGE STANDARD; PRT; 1024 AA.
AC P47321; Q49190;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG075.
GN MG075.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 1-156; 269-402; 643-736 AND 808-947 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U39688; AAC71293.1; -.
DR EMBL; U01715; AAC43189.1; ALT_INIT.
DR EMBL; U02251; AAD12514.1; -.
DR EMBL; U01749; AAD10562.1; -.
DR EMBL; U01775; AAD10595.1; -.
DR TIGR; MG075; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 951 971 POTENTIAL.
SQ SEQUENCE 1024 AA; 116424 MW; 5B1540E6855CB554 CRC64;

Query Match 8.1%; Score 89; DB 1; Length 1024;

Best Local Similarity 23.9%; Pred. No. 9.2;
Matches 47; Conservative 30; Mismatches 74; Indels 46; Gaps 9;

```
Qy         26 QFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP---YFLFD-----------TIYQG 70

Db        301 QFGEHKNGQPLSLLKVLSGLNNEFKQLFNYV--RPGLGDFVSDLIQSSSQSSNKKTVYQL 358

Qy         71 LGIDRIAACYTIEDGVVVDAGSAI-TIDIISNSIHLGGFILPGIANYKKIYSHISPRLKS 129

Db        359 LFENKTTIIHLLQDLNISELNSVLPVVDILFEGINSAESLYQRIQSFKDL---IVPALKA 415

Qy        130 EFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYF---------TGGDGQFLA 180

Db        416 DKQLK-SLEAIILAVLDNPNTYVFDLVY-------QNKSILFNLLSDFLKNTANTLPFLQ 467

Qy        181 NYFD--------HAIYD 189

Db        468 EQFDIVNHLFANEAIFD 484
```

RESULT 7
GREA_BORBU
ID GREA_BORBU STANDARD; PRT; 901 AA.
AC O51157;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor
DE greA).
GN GREA OR BB0132.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi.";
RL Nature 390:580-586(1997).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
CC ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
CC SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
CC OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
CC LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CC CLEAVAGE FACTORS SUCH AS GREA OR GREB ALLOWS THE RESUMPTION OF
CC ELONGATION FROM THE NEW 3'TERMINUS. GREA RELEASES SEQUENCES OF
CC 2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 GREA ASSOCIATED DOMAIN 1 (GRAD1).
CC -!- SIMILARITY: CONTAINS 1 GREA ASSOCIATED DOMAIN 2 (GRAD2).
CC -!- SIMILARITY: BELONGS TO THE GREA/GREB FAMILY.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; AE001124; AAC66506.1; -.
DR HSSP; P21346; 1GRJ.
DR TIGR; BB0132; -.
DR InterPro; IPR001437; GreA_GreB.
DR Pfam; PF01272; GreA_GreB; 1.

```
DR   Pfam; PF03449; GreA_GreB_N; 1.
DR   ProDom; PD004918; GreA_GreB; 1.
DR   PROSITE; PS00829; GREAB_1; 1.
DR   PROSITE; PS00830; GREAB_2; 1.
KW   Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT   DOMAIN       20    170       GRAD2.
FT   DOMAIN      171    681       GRAD1.
FT   DOMAIN      741    901       GREA.
FT   DOMAIN      784    815       COILED COIL (POTENTIAL).
SQ   SEQUENCE   901 AA;  105314 MW;  FBB81088D735A5F9 CRC64;

  Query Match             8.1%;  Score 88.5;  DB 1;  Length 901;
  Best Local Similarity   19.7%;  Pred. No. 8.8;
  Matches   42;  Conservative   43;  Mismatches   65;  Indels   63;  Gaps   10;

Qy          4 CDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIF--YINV-----NEHLKEHLKNQKNFI 56
              | | |  :: |::  |    |::  :|    :||  | ||     : |::| : |
Db        439 CYIYNERSSNLNEKMYDKFKIEK--DFYKRYEIFINYCNVKGIVKDLHIEEEEMLN----- 491

Qy         57 NLEPYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANY 116
                |||                    : |:  : :   ::|: :   . ||  : |:
Db        492 ----YFL--------------------IYVNNFTKVDHHVVSSYV-----ILKSLGNF 520

Qy        117 KKIYSHISPRLKSEFNTQVSLDAFPQKTMD----ALSYGVFKGIYLLIKDAAQNKKLYFT 172
               :  |     :: : | :  |  :|:  :|     |:  : | : :|||:   |  ||:
Db        521 SEELSS-KVNIEKDVNLESLLKEYPKGIVDLFDSILNAEIKKELVVLIKEELSNWILYY- 578

Qy        173 GGDGQFLANYFDHAIYDKLLIFRGMKKIIKENP 205
                      | |::  :|    :: : ||:|
Db        579 -------KELFPHSVNKRL-----VESLYKEDP 599


RESULT 8
COAA_RHILO
ID   COAA_RHILO     STANDARD;      PRT;   319 AA.
AC   Q98CS9;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Pantothenate kinase (EC 2.7.1.33) (Pantothenic acid kinase).
GN   COAA OR MLR5019.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
CC   -!- CATALYTIC ACTIVITY: ATP + pantothenate = ADP + D-4'-
CC       phosphopantothenate.
CC   -!- PATHWAY: Coenzyme A (CoA) biosynthesis; first step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC   -!- SIMILARITY: BELONGS TO THE PANTOTHENATE KINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP003005; BAB51542.1; -.
DR   InterPro; IPR004566; PanK_bact.
DR   TIGRFAMs; TIGR00554; panK_bact; 1.
KW   Transferase; Kinase; ATP-binding; Coenzyme A biosynthesis;
KW   Complete proteome.
FT   NP_BIND      97    104       ATP (POTENTIAL).
SQ   SEQUENCE   319 AA;  36661 MW;  306F394E7EF9F2FE CRC64;

  Query Match             8.0%;  Score 87;  DB 1;  Length 319;
  Best Local Similarity   20.6%;  Pred. No. 3.6;
  Matches   51;  Conservative   43;  Mismatches   78;  Indels   76;  Gaps   11;

Qy         21 TLNIDQFLEFKN--------EQKIFYINVNEHLKEHLK-NQKNFINLEPYFLFDTIYQGL 71
              ||: |:|   ::        | |  |::::  |  |:: :|  |   : :|
Db         31 TLSEDEFRRLRSLNDPVDLEEVKRIYLSLSRLLSAHVEASQLLFRQRQAFF--------N 82

Qy         72 GIDRIAACYTIEDGVVVDAGSAIT---------------IDIISNSIHLGGFILPG--- 112
               :| :   : |     |  | : |                :|:|:     ||:||
Db         83 AVDIVKTPFIIGIAGSVAVGKSTTARVLKELLARWPSSPKVDLITTD----GFLLPNEVL 138

Qy        113 ---------------IANYKKIYSHIS---PRLKSEFNTQVSLDAFPQK--TMDALSYG 151
                            :   :  | |    | :::   : :: |  | :  |:|
Db        139 RRENLMERKGFPDSYDVGALLRFLSGIKSALPDVRAPVYSHLTYDVIPGEFVTIDRPDIL 198

Qy        152 VFKGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIY---DKLLI-------FRGMKKII 201
              :|:|| :|     |  ||   |    ||:::|| |||   |: ||       |  :::
Db        199 IFEGINVL-----QPGKLPQDGKIVPFLSDFFDFAIYIDADEKLIHEWYISRFMRLRETA 253

Qy        202 KENPNLLY 209
                ||:  :
Db        254 FRNPDSFF 261


RESULT 9
MQO_CAMJE
ID   MQO_CAMJE      STANDARD;      PRT;   448 AA.
AC   P56954;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable malate:quinone oxidoreductase (EC 1.1.99.16) (Malate
DE   dehydrogenase [acceptor]) (MQO).
GN   MQO OR CJ0393C.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC   Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
CC   -!- FUNCTION: CATALYZES OXIDATION OF MALATE TO OXALOACETATE IN THE
CC       CITRIC ACID CYCLE. DONATES ELECTRONS TO QUINONES OF THE ELECTRON
CC       TRANSFER CHAIN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: (S)-malate + acceptor = oxaloacetate + reduced
CC       acceptor.
CC   -!- COFACTOR: FAD; TIGHTLY BOUND (BY SIMILARITY).
CC   -!- PATHWAY: Tricarboxylic acid cycle.
CC   -!- SIMILARITY: BELONGS TO THE MQO FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AL139075; CAB74229.1; -.
KW   Oxidoreductase; Tricarboxylic acid cycle; Flavoprotein; FAD;
KW   Complete proteome.
SQ   SEQUENCE   448 AA;  50586 MW;  66D54F010F30A025 CRC64;

  Query Match             7.9%;  Score 86.5;  DB 1;  Length 448;
  Best Local Similarity   20.0%;  Pred. No. 5.8;
  Matches   46;  Conservative   44;  Mismatches   81;  Indels   59;  Gaps   14;

Qy          5 DIGNSNANFLDD------NKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINL 58
              | |  : | ::       | :   |  : :  : :  || :  : | : | |:    :|
Db        170 DFGKMSINLIEQAQKQNKNTFVAFN-QEIIHIEKKDDIFILKTSNHQEYHAKSV--VVNA 226

Qy         59 EPYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITI--DIISNSIHLGGFILPGIANY 116
                :|:      ||:|:  :|: :       |||       :::  :::       : |
Db        227 GAHSLYLAHKMNLGMDK--SCWPV-------AGSFYLTKQKLLNGKVYM-------VQNP 270

Qy        117 KKIYS--HISPRLKSEFNTQV-------------SLDAFPQKTMDALSYG--VFKGIYL 158
              |  ::  |  | | :: ||:               | : |:  :||     | |  :
Db        271 KLPFAALHGDPDLLADMNTRFGPTALVIPKLERYHGLKSVPE-FFEALKLDKTVLKVTFN 329

Qy        159 LIKDAAQNKKLYFTGGDGQFLANY-FDHAIYDKLLIFRGMKKIIKENPNL 207
              : |||      :::         || |:    || |  :  |||:   |:|
Db        330 MFKDATIRNYIFY---------NYLFELPFVDKSLFVKDAKKIV---PSL 367


RESULT 10
SSP5_STRGN
ID   SSP5_STRGN     STANDARD;      PRT;  1500 AA.
AC   P16952; Q54184;
DT   01-AUG-1990 (Rel. 15, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Agglutinin receptor precursor (SSP-5).
GN   SSP5 OR SSPB.
OS   Streptococcus gordonii.
OC   Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC   Streptococcus.
OX   NCBI_TaxID=1302;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M5;
RX   MEDLINE=90236997; PubMed=2185241;
RA   Demuth D.R., Golub E.E., Malamud D.;
RT   "Streptococcal-host interactions. Structural and functional analysis
RT   of a Streptococcus sanguis receptor for a human salivary
RT   glycoprotein.";
RL   J. Biol. Chem. 265:7120-7126(1990).
RN   [2]
RP   REVISIONS, SEQUENCE FROM N.A.
RC   STRAIN=M5;
RX   MEDLINE=96310377; PubMed=8733238;
RA   Demuth D.R., Duan Y., Brooks W., Holmes A.R., McNab R.,
RA   Jenkinson H.F.;
RT   "Tandem genes encode cell-surface polypeptides SspA and SspB which
RT   mediate adhesion of the oral bacterium Streptococcus gordonii to
RT   human and bacterial receptors.";
RL   Mol. Microbiol. 20:403-413(1996).
CC   -!- FUNCTION: MAY BIND SIALIC ACID RESIDUES OF SALIVARY AGGLUTININ
CC       (SAG) IN A CALCIUM-DEPENDENT REACTION. THE INTERACTION OF SAG WITH
CC       ITS RECEPTOR IN VARIOUS ORAL STREPTOCOCCI MODULATE BACTERIAL
CC       COLONIZATION OF ORAL TISSUE AND IS ASSOCIATED WITH REDUCED LEVELS
CC       OF DENTAL CARIES.
CC   -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC       an amide bond (Potential).
CC   -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U40026; AAC44100.1; -.
DR   PIR; A35186; A35186.
DR   InterPro; IPR001899; Gram_pos_anchor.
DR   Pfam; PF00746; Gram_pos_anchor; 1.
DR   ProDom; PD153432; Csurface_antigen; 2.
DR   TIGRFAMs; TIGR01167; LPXTG_anchor; 1.
DR   PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
KW   Cell wall; Peptidoglycan-anchor; Calcium-binding; Repeat; Signal.
FT   SIGNAL       1     38
FT   CHAIN       39   1470       AGGLUTININ RECEPTOR.
FT   PROPEP    1471   1500       REMOVED BY SORTASE (POTENTIAL).
FT   SIMILAR    164    470       TO M PROTEIN OF S.PYOGENES.
FT   DOMAIN     164    470       4 X APPROXIMATE TANDEM REPEATS, HR1.
FT   REPEAT     164    241       1.
FT   REPEAT     242    323       2.
FT   REPEAT     324    405       3.
FT   REPEAT     406    470       4.
FT   DOMAIN     771    887       3 X APPROXIMATE TANDEM REPEATS, PR1.
FT   DOMAIN    1414   1436       PRO-RICH (PR2).
FT   CA_BIND    220    235       POTENTIAL.
FT   CA_BIND    301    316       POTENTIAL.
FT   CA_BIND    931    950       POTENTIAL.
FT   CA_BIND   1300   1315       POTENTIAL.
FT   SITE      1467   1471       LPXTG SORTING SIGNAL (POTENTIAL).
FT   MOD_RES   1470   1470       AMIDE-LINKED TO CELL WALL (POTENTIAL).
SQ   SEQUENCE   1500 AA;  164552 MW;  DCF190E7D44D889F CRC64;

  Query Match             7.8%;  Score 85.5;  DB 1;  Length 1500;
  Best Local Similarity   24.0%;  Pred. No. 28;
  Matches   42;  Conservative   27;  Mismatches   69;  Indels   37;  Gaps    8;

Qy          5 DIGN--------SNANFLDDNKYFTLNIDQFLEFKNEQKI-FYINVNEHLKEHLKNQKNF 55
              | ||         :|   :   |  |:::|   |  : | :  |   |  |   |  | :
Db       1238 DFGNGYATEVVVNNVPKITPKKDVTVSLDPTSENLDGQTVQLYQTFNYRLIGGLIPQNHS 1297

Qy         56 INLEPYFLFDTIYQGLGIDRIAACY--------TIEDGVVVDAGSAITIDIISNSIHLGG 107
               || |   |  |.  | |:     |         |::|| |: ||: :|    : :    |
Db       1298 EELEDYSFVDD-YDQAG-DQYTGNYKTFSSLNLTMKDGSVIKAGTDLTSQTTAETDATNG 1355

Qy        108 FILPGIANYKKIYSHISPRLKSEFNTQVSLDA-FPQKT---MDALSYGVFKGIYL 158
                          :: | | :|  ::|||: |  :|   |  :: | |:  |:
Db       1356 I-------------VTVRFKEDFLQKISLDSPFQAETYLQMRRIAIGTFENTYV 1396


RESULT 11
TRME_MYCGE
ID   TRME_MYCGE     STANDARD;      PRT;   442 AA.
AC   P47254; Q49330;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable tRNA modification GTPase trmE.
GN   TRME OR THDF OR MG008.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 1-88 FROM N.A.
```

```
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- FUNCTION: EXHIBITS A VERY HIGH INTRINSIC GTPASE HYDROLYSIS RATE.
CC       INVOLVED IN THE BIOSYNTHESIS OF THE HYPERMODIFIED NUCLEOSIDE 5-
CC       METHYLAMINOMETHYL-2-THIOURIDINE, WHICH IS FOUND IN THE WOBBLE
CC       POSITION OF SOME TRNAS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING
CC       PROTEINS. TRME SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39679; AAC71224.1; -.
DR   EMBL; U02216; AAD12512.1; ALT_INIT.
DR   TIGR; MG008; -.
DR   InterPro; IPR005289; GTP-bindding_dom.
DR   InterPro; IPR002917; MMR_HSR1.
DR   InterPro; IPR005225; Small_GTP.
DR   InterPro; IPR004520; ThdF.
DR   Pfam; PF01926; MMR_HSR1; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   TIGRFAMs; TIGR00450; thdF; 1.
DR   TIGRFAMs; TIGR00650; MG442; 1.
KW   tRNA processing; GTP-binding; Complete proteome.
FT   NP_BIND     224    231       GTP (POTENTIAL).
FT   NP_BIND     271    275       GTP (POTENTIAL).
FT   NP_BIND     331    334       GTP (POTENTIAL).
SQ   SEQUENCE   442 AA;  50806 MW;  C6D50D3274B0EB32 CRC64;

  Query Match             7.7%;  Score 84.5;  DB 1;  Length 442;
  Best Local Similarity   23.8%;  Pred. No. 8.3;
  Matches   49;  Conservative   37;  Mismatches   81;  Indels   39;  Gaps   11;

Qy         22 LNIDQFLEFKNEQ-KIFYIN-----VNEHLKEHLKNQKNFINLEPYFLFDTIYQ-GLGID 74
              :||| : |: :|| ::  :|     : | ||  ::| |    |   |    | :  :|
Db        173 VNID-YPEYLDEQVELSTLNNKVKLIIEKLKRIIENSKQLKKLHDPFKIAIIGETNVGKS 231

Qy         75 RIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFIL-----PGIANYKKIYSHISPRLKS 129
                :      :|  :|      | |::    :| |:::      ||  :|        :
Db        232 SLLNALLNQDKAIVSNIKGSTRDVVEGDFNLNGYLIKILDTAGIRKHKSGLEKAGIKKSF 291

Qy        130 EFNTQVS-----LDA-FPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFT-GGDGQFLANY 182
              |   | :     |||  |:| :: :|:  ||          :||| :|       : |
Db        292 ESIKQANLVIYLLDATHPKKDLELISF--FK----------KNKKDFFVFYNKKDLITNK 339

Qy        183 FDHAIYDKLLIFRGMKKIIKENPNLL 208
              |:::|  |       :| |||  :||
Db        340 FENSISAK-------QKDIKELVDLL 358


RESULT 12
FLIR_BORBU
ID   FLIR_BORBU     STANDARD;      PRT;   269 AA.
AC   Q44907; Q44761;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Flagellar biosynthetic protein fliR.
GN   FLIR OR BB0273.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RA   Dunn J.J., Butler-Loffredo L., Kieleczawa J., Medalle J., Luft B.J.;
RL   Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=212;
RA   Ge Y., Charon N.W.;
RL   Submitted (FEB-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K.; Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: ROLE IN FLAGELLAR BIOSYNTHESIS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE FLIR/MOPE/SPAR FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U43739; AAA85597.1; -.
DR   EMBL; L75945; AAB58970.1; -.
DR   EMBL; AE001137; AAC66675.1; -.
DR   TIGR; BB0273; -.
DR   InterPro; IPR002010; Bac_export_1.
DR   Pfam; PF01311; Bac_export_1; 1.
KW   Flagella; Transmembrane; Complete proteome.
FT   TRANSMEM     11     31       POTENTIAL.
FT   TRANSMEM     36     56       POTENTIAL.
FT   TRANSMEM     75     95       POTENTIAL.
FT   TRANSMEM    127    147       POTENTIAL.
FT   TRANSMEM    190    210       POTENTIAL.
FT   TRANSMEM    219    239       POTENTIAL.
SQ   SEQUENCE   269 AA;  30876 MW;  EAA365C766CC88FE CRC64;

  Query Match             7.6%;  Score 83;  DB 1;  Length 269;
  Best Local Similarity   22.9%;  Pred. No. 6.3;
  Matches   48;  Conservative   27;  Mismatches   55;  Indels   80;  Gaps   11;

Qy         10 NANFLDDNKYFTL-----NIDQFLEFK---NEQKIFYINVNEHLKEHLKNQKNFINLEPY 61
                | |||   | ||:      |  ||:|    :  || | |
Db          9 NLNFL-VLKSFTILPVLVRIFMFLKFSPFFSTIKIGYFN---------------------- 46

Qy         62 FLFDTIYQGLGIDRIAACYTIEDGV---VVDAGSAIT-------IDIISNSIHLGGFILP 111
              | |  |   : :::|   | ::: :   ::  | ||        ::|| |  || ||
Db         47 FFFSLILSVIVVEKIKIIYPLDNMLSFALILLGEAILGLIQAFFVNIIFNVFHLVGFF-- 104

Qy        112 GIANYKKIYSHISPRLKSEFNTQVSL------DAFPQKTMDALSYGVFKGIYLLIKDAAQ 165
                                |: |: |      | | ::    :|  :|  ::||:
Db        105 ------------------FSNQIGLAYANIFDVFSEEDSMIISQ-IFAYLFLLL----- 139

Qy        166 NKKLYFTGGDGQFLANYFDHAIYDKLLIFR 195
                  |   |  ||  :|   |:| :|  |
Db        140 -----FLSSD--FLLRFFVIGIHDSVLNIR 162


RESULT 13
PPNK_BORBU
```

```
ID   PPNK_BORBU     STANDARD;      PRT;   279 AA.
AC   O51291;
DT   30-MAY-2000 (Rel. 39, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
DE   (Poly(P)/ATP NAD kinase).
GN   PPNK OR BB0311.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
CC   -!- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC       ATP and other nucleoside triphosphates as well as inorganic
CC       polyphosphate as a source of phosphorus (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
CC   -!- COFACTOR: Requires divalent metal ions for activity (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001138; AAC66699.1; ALT_INIT.
DR   TIGR; BB0311; -.
DR   InterPro; IPR002504; ATP_NADK.
DR   Pfam; PF01513; NAD_kinase; 1.
KW   Transferase; Kinase; NAD; NADP; Complete proteome.
SQ   SEQUENCE   279 AA;  30966 MW;  1592FC198FAFE96C CRC64;

  Query Match             7.5%;  Score 82.5;  DB 1;  Length 279;
  Best Local Similarity   22.4%;  Pred. No. 7.2;
  Matches   47;  Conservative   39;  Mismatches   93;  Indels   31;  Gaps    9;

Qy          3 LCDIGNSNANFLDDNKY--FTLNIDQFLEFK---NEQKIFYINVNEHLKEHL-KNQKNFI 56
              :  |   |  || | |   |   ||:|      |:: : :: |::| |: : |   | |
Db         81 IISINMGNVGFLADIKIEDFKKVIDRFFNNSLVINKKFLLHVTVSQHGKDLISKYALNDI 140

Qy         57 NLEPYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILP----- 111
                :    |  ||  | ::  :      ||::|   :  |      |   || ||
Db        141 IIRSSVLNKMIYVDLMVNSESFLSYKSDGIIVSTPTGST----GYSFSAGGPILEADLEG 196

Qy        112 ---------GIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKD 162
                      : |   ::| :| :|   |: :  : |     :| :::| | |: :: :
Db        197 FXLTPISPHSVYNRSFVFSKLS-KLSISFSKEYFI-AAASIFLDGINFGSF-GVDVVFEF 253

Qy        163 AAQNKKLYFTGGDGQFLANYFDHAIYDKLL 192
                :: | |     |  : |   : :|||
Db        254 KISSQSLNFV----SFCTDTFVKRLKNKLL 279


RESULT 14
PTA_BUCAI
```

```
ID   PTA_BUCAI      STANDARD;      PRT;   708 AA.
AC   P57273;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phosphate acetyltransferase (EC 2.3.1.8) (Phosphotransacetylase).
GN   PTA OR BU176.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tokyo 1998;
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- CATALYTIC ACTIVITY: Acetyl-CoA + phosphate = CoA + acetyl
CC       phosphate.
CC   -!- PATHWAY: LAST OF TWO STEPS IN THE CONVERSION OF ACETATE TO
CC       ACETYL-COA.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; CONTAINS A DOMAIN WITH SOME
CC       SIMILARITIES WITH COBQ/BIOD.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE PHOSPHATE
CC       ACETYLTRANSFERASE AND BUTYRYLTRANSFERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP001118; BAB12893.1; -.
DR   InterPro; IPR002505; PTA_PTB.
DR   InterPro; IPR004614; Pta.
DR   Pfam; PF01515; PTA_PTB; 1.
DR   TIGRFAMs; TIGR00651; pta; 1.
KW   Transferase; Acyltransferase; Complete proteome.
FT   DOMAIN      388    708       PHOSPHATE ACETYLTRANSFERASE.
SQ   SEQUENCE   708 AA;  80770 MW;  1EDC4394824ED89A CRC64;

  Query Match             7.5%;  Score 82.5;  DB 1;  Length 708;
  Best Local Similarity   22.9%;  Pred. No. 21;
  Matches   56;  Conservative   41;  Mismatches   69;  Indels   79;  Gaps   15;

Qy         10 NANFLDDNKY-FTLNIDQFLEFKNEQKIF------------------------------ 37
              |: || :||| |   :    : ||| |
Db        174 NSPFL-ENKYDFIKKLIVLKKIKNETKTIVPKKILKNNFFSIIACIPWNRNIVTTRVIDL 232

Qy         38 --YINVNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAIT 95
                 ::|: :|   :|  :|| | :|   :||| :  | :::    :::   |:|   |
Db        233 FNFLNI-QH--TNLVQKKNHI-IEEIIIFDTHHLNL-LNK----HSLNTLVIV---SFSR 280

Qy         96 IDIISNSIH-------LGGFILPGIANYKK----IYSHISPRLKSEFNTQVSLDAFPQKT 144
              :|:  | ::        :   || ||   ||   :|  :  |  | | |:       : |
Db        281 VDVFLNVLNCNVNRSKVKCIILTGILKLKKNIASLYKFLIKRSISIFFTE-------KNT 333

Qy        145 MDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKEN 204
              :: ||        : :||    |||       ::::|:| |:       |   ||  | |
Db        334 IEILSQLQNFNFDISVKDITYIKKLQ------RYISNFFCHS------SFMFFKK--KYN 379

Qy        205 PNLLY 209
               |::|
Db        380 INVIY 384


RESULT 15
YBFT_BACSU
```

```
ID   YBFT_BACSU     STANDARD;      PRT;   249 AA.
AC   O31458;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ybfT.
GN   YBFT.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Haga K., Liu H., Yasumoto K., Takahashi H., Yoshikawa H.;
RT   "Sequence analysis of the 70kb region between 17 and 23 degree of the
RT   Bacillus subtilis chromosome.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: BELONGS TO THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE
CC       ISOMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB006424; BAA33133.1; -.
DR   EMBL; Z99105; CAB12030.1; -.
DR   HSSP; P09375; 1DEA.
DR   SubtiList; BG12746; ybfT.
DR   InterPro; IPR000457; Glucosamine_iso.
DR   InterPro; IPR004547; NagB.
DR   Pfam; PF01182; Glucosamine_iso; 1.
DR   TIGRFAMs; TIGR00502; nagB; 1.
DR   PROSITE; PS01161; GLC_GALNAC_ISOMERASE; 1.
KW   Hypothetical protein; Isomerase; Complete proteome.
FT   ACT_SITE     67     67       BY SIMILARITY.
SQ   SEQUENCE   249 AA;  27289 MW;  386A7D7301C420D3 CRC64;

  Query Match             7.5%;  Score 82;  DB 1;  Length 249;
  Best Local Similarity   21.1%;  Pred. No. 6.9;
  Matches   43;  Conservative   31;  Mismatches   74;  Indels   56;  Gaps    6;

Qy          6 IGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYIN-VNEHLKEHLKNQKNFINLEPYFLF 64
              | :  |  :| :|  | |:|::          | : ::||| :|:  | : |:
Db         48 ISDYQAGEIDFSKVTTFNLDEYAGLSPSHPQSYNHFMHEHLFQHINMQPDHIH------- 100

Qy         65 DTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHIS 124
                | ||      |||   || :    |  : |  |  : |: ||  ||
Db        101 --IPQGDNPQLEAACKVYEDLIRQAGGIDVQILGIGANGHI-GFNEPG------------ 145

Qy        125 PRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLANYFD 184
                |:|  :  :    : |: |                          | ||         |
Db        146 ----SDFEDRTRVVKLSESTIQA-------------------NARFFGG---------D 172

Qy        185 HAIYDKLLIFRGMKKIIKENPNLL 208
              :  :| |  |:| |:: : :::
Db        173 PVLVPRLAISMGIKTIMEFSKHIV 196


Search completed: June 24, 2003, 21:51:29
Job time : 5.63509 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 17.0461 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-21
Perfect score: 1093
Sequence: 1 MLLCDIGNSNANFLDDNKYF.....KLLIFRGMKKIIKENPNLLY 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1093	100.0	209	16	Q9PIA9	Q9pia9 campylobact
2	391.5	35.8	223	16	O25533	O25533 helicobacte
3	384.5	35.2	223	16	Q9ZKY6	Q9zky6 helicobacte
4	177.5	16.2	229	16	O67753	O67753 aquifex aeo
5	141	12.9	256	16	Q8RFE4	Q8rfe4 fusobacteri
6	139.5	12.8	246	16	Q9WZY5	Q9wzy5 thermotoga
7	132	12.1	276	16	Q8YQD7	Q8yqd7 anabaena sp
8	126.5	11.6	224	16	Q98Q93	Q98q93 mycoplasma
9	121.5	11.1	262	16	O51477	O51477 borrelia bu
10	120.5	11.0	248	16	Q9HWC1	Q9hwc1 pseudomonas
11	116.5	10.7	257	16	P74045	P74045 synechocyst
12	111	10.2	261	16	Q9A6Z1	Q9a6z1 caulobacter
13	110	10.1	254	16	Q9KGH5	Q9kgh5 bacillus ha
14	105	9.6	3973	5	O96204	O96204 plasmodium
15	104	9.5	479	16	Q8R663	Q8r663 fusobacteri
16	103	9.4	866	16	Q8XIW9	Q8xiw9 clostridium
17	102.5	9.4	967	5	O77305	O77305 plasmodium
18	102.5	9.4	967	5	Q8WP96	Q8wp96 plasmodium
19	98.5	9.0	356	2	Q9AKH2	Q9akh2 rickettsia
20	98.5	9.0	3119	5	Q25857	Q25857 plasmodium
21	98	9.0	259	16	Q8YAC5	Q8yac5 listeria mo
22	97	8.9	967	16	O26050	O26050 helicobacte
23	95.5	8.7	662	5	P90663	P90663 aedes aegyp
24	95	8.7	969	16	Q9ZJA3	Q9zja3 helicobacte
25	94.5	8.6	274	16	Q9CD56	Q9cd56 mycobacteri
26	94.5	8.6	481	16	Q9K691	Q9k691 bacillus ha
27	94	8.6	465	16	Q9PJ81	Q9pj81 campylobact
28	93.5	8.6	381	12	Q9YVM5	Q9yvm5 melanoplus
29	93	8.5	259	16	Q92F54	Q92f54 listeria in
30	92.5	8.5	156	10	Q43669	Q43669 trollius le
31	92.5	8.5	2155	12	Q64958	Q64958 avian infec
32	92.5	8.5	6629	12	Q91QT2	Q91qt2 avian infec
33	92	8.4	1383	2	Q9K5L9	Q9k519 anabaena sp
34	91.5	8.4	272	16	O06282	O06282 mycobacteri
35	91.5	8.4	356	2	Q9AKM7	Q9akm7 rickettsia
36	90.5	8.3	804	5	Q8WRL8	Q8wrl8 tetrahymena
37	90	8.2	1447	5	Q8T1Y0	Q8tly0 dictyosteli
38	89.5	8.2	242	16	Q9PCI4	Q9pci4 xylella fas
39	89.5	8.2	256	16	Q97RF5	Q97rf5 streptococc
40	89.5	8.2	391	16	Q8R651	Q8r651 fusobacteri
41	89.5	8.2	2340	5	O97298	O97298 plasmodium
42	89	8.1	259	16	Q8XHL5	Q8xhl5 clostridium
43	89	8.1	566	16	Q8R624	Q8r624 fusobacteri
44	88.5	8.1	431	10	Q9SA92	Q9sa92 arabidopsis
45	88.5	8.1	531	5	Q966P6	Q966p6 caenorhabdi

ALIGNMENTS

RESULT 1
Q9PIA9
ID Q9PIA9 PRELIMINARY; PRT; 209 AA.
AC Q9PIA9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Cj0394c.
GN CJ0394C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139075; CAB74230.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 209 AA; 23935 MW; C49313AA9E3D2D19 CRC64;

Query Match 100.0%; Score 1093; DB 16; Length 209;
Best Local Similarity 100.0%; Pred. No. 2.3e-82;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60
```

```
Qy          61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120

Qy         121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180

Qy         181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
               |||||||||||||||||||||||||||||
Db         181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209


RESULT 2
O25533
ID   O25533      PRELIMINARY;      PRT;   223 AA.
AC   O25533;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein HP0862.
GN   HP0862.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000596; AAD07916.1; -.
DR   TIGR; HP0862; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   223 AA;  24655 MW;  C1D13E8BEC1EB9AF CRC64;

  Query Match             35.8%;  Score 391.5;  DB 16;  Length 223;
  Best Local Similarity   39.2%;  Pred. No. 1.4e-24;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 3
Q9ZKY6
ID   Q9ZKY6      PRELIMINARY;      PRT;   223 AA.
AC   Q9ZKY6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PUTATIVE.
GN   JHP0796.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001509; AAD06372.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   223 AA;  24754 MW;  B31377DA1AEF6E0E CRC64;

  Query Match             35.2%;  Score 384.5;  DB 16;  Length 223;
  Best Local Similarity   38.2%;  Pred. No. 5.2e-24;
  Matches   78;  Conservative   46;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LILCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYIGLGIDRQMACLAVVNGVIVDAGSAITIDLVKEGKHLGGCILPGLAQYVHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |:  |    ||:  |: | ||::||:   :|  |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCIQHLAKDQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 4
O67753
ID   O67753      PRELIMINARY;      PRT;   229 AA.
AC   O67753;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein AQ_1924.
GN   AQ_1924.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
DR   EMBL; AE000763; AAC07720.1; -.
```

```
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   229 AA;  26068 MW;  64C347A285FABE9A CRC64;

  Query Match           16.2%;  Score 177.5;  DB 16;  Length 229;
  Best Local Similarity  25.8%;  Pred. No. 5.9e-07;
  Matches   62;  Conservative   44;  Mismatches   83;  Indels   51;  Gaps   10;

Qy          2 LLCDIGNSNANF-LDDNKYFTLNIDQFLEFKNEQ---------KIFYINVNEHLKEHLK- 50
              |  |:|||: :  | : |     :  ||:  :|:         |   |:| :   | ::
Db          4 LTVDVGNSSVDIALWEGK----KVKDFLKLSHEEFLKEEFPKLKALGISVKQSFSEKVRG 59

Qy         51 --------NQKNFINLEPYFLFDTIYQGLGIDRIAACYTI-----EDGVVVDAGSAITID 97
                    ::||     |   :|    || ||:|  |:      :: ||: ||:|: ||
Db         60 KIPKIKFLKKENFPIQVDYKTPET----LGTDRVALAYSAKKFYGKNVVVISAGTALVID 115

Qy         98 IISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIY 157
              ::      ||||  |:    || | ::  :   |  :| :    : | : :  | ::
Db        116 LVLEGKFKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEI-FLGRSTRECVLGGAYREST 174

Qy        158 LLIKDA--------AQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
                ||           :  |:  |||:|:: : :    ||| ||: |||:       ||||
Db        175 EFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGMR-------NLLY 224


RESULT 5
Q8RFE4
ID   Q8RFE4      PRELIMINARY;      PRT;   256 AA.
AC   Q8RFE4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bvg accessory factor.
GN   FN0761.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010586; AAL94957.1; -.
KW   Complete proteome.
SQ   SEQUENCE   256 AA;  28601 MW;  CDF681127F08914B CRC64;

  Query Match           12.9%;  Score 141;  DB 16;  Length 256;
  Best Local Similarity  25.6%;  Pred. No. 0.00069;
  Matches   65;  Conservative   46;  Mismatches   85;  Indels   58;  Gaps   16;

Qy          5 DIGNSN--ANFLDDN-----------------KYFTL--NIDQF--LEFKNEQKIFYIN 40
              ||||::      |:|                  :||:   || ::  :  |    |   :
Db          6 DIGNTHIVTGIYDNNGELISTFRIATNDKMTEDEYFSYFNNITKYNEISIKKVDAILISS 65

Qy         41 VNEHL--------KEHLKNQKNFINLE---PY-FLFDTIYQGLGIDRIAAC------YTI 82
              |  ::        ::: | :   ::||   |: |     | | | |||          |
Db         66 VVPNIIITFQFFARKYFKVEATIVDLEKKLPFTFAKGINYTGFGADRIIDITEAMQKYPD 125

Qy         83 EDGVVVDAGSAITIDIISNSIHLGGFILPGI-ANYKKIYSHIS--PRLKSEFNTQVSLDA 139
              :: |: | |:| | |::   :::|| |||||  :   :| : :  ||:|  | |  |:
Db        126 KNLVIFDFGTATTYDVLKKGVYIGGGILPGIDMSINALYGNTAKLPRVK--FTTPSSV-- 181

Qy        140 FPQKTMDALSYGVFKG----IYLLIKDAAQ--NKKLYF--TGGDGQFLANYFDHA-IYDK 190
                  ||  :   :| |    |  :||   :  |::::   ||| |: |:   '    ||
Db        182 LGTDTMKQIQAAIFFGYAGQIKHIIKKINEELNEEIFVLATGGLGKILSAEIDEIDEYDA 241

Qy        191 LLIFRGMKKIIKEN 204
                |  :|:  : | |
Db        242 NLSLKGLYTLYKLN 255


RESULT 6
Q9WZY5
ID   Q9WZY5      PRELIMINARY;      PRT;   246 AA.
AC   Q9WZY5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TM0883.
GN   TM0883.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001754; AAD35964.1; -.
DR   TIGR; TM0883; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   246 AA;  27154 MW;  9E0309AD462CF266 CRC64;

  Query Match           12.8%;  Score 139.5;  DB 16;  Length 246;
  Best Local Similarity  22.7%;  Pred. No. 0.00087;
  Matches   58;  Conservative   38;  Mismatches   82;  Indels   77;  Gaps   10;

Qy          2 LLCDIGNSNANF--LDDNKYF--------------------TLNIDQFLEFKN------ 32
              || |:||::: |   :| | |                      |  |   | |
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         33 -----------EQKIFYIN------VNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDR 75
                        || |:|:       |  :| ::||                  :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNVKNPSE---------------VGADR 106

Qy         76 IAACYTI-----EDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSE 130
              :|          ::|:::| |:| |:|:: |  : || ||||   :  ::|      |
Db        107 VANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGF--FMMVHSLFRGTAKLP 164

Qy        131 FNTQVSLDAFPQK-TMDALSYGVFKG-IYLL------IKDAAQNKKLYFTGGDGQFLANY 182
                    |    | | : :  ||  | :| |      ||:   :  :  |||  : : :
Db        165 LVEVKPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDM 224

Qy        183 FDHAIYDKLLIFRGM 197
                | |:|: |  :|:
Db        225 IKHEIFDEDLTIKGV 239


RESULT 7
Q8YQD7
ID   Q8YQD7      PRELIMINARY;      PRT;   276 AA.
AC   Q8YQD7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Alr3896.
```

```
GN   ALR3896.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003594; BAB75595.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   276 AA;  30272 MW;  2AD04CA693E56F25 CRC64;

  Query Match             12.1%;  Score 132;  DB 16;  Length 276;
  Best Local Similarity   25.8%;  Pred. No. 0.0041;
  Matches   69;  Conservative   38;  Mismatches   84;  Indels   76;  Gaps   14;

Qy          2 LLCDIGNSN---ANFLDDNKYFTLNIDQF--------------LEFKNEQK-IFY----- 38
              |  :||||    | |: ::  || : :                ||  :|:| ||:
Db         15 LALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEVGSEEKEIFFTFFPL 74

Qy         39 --------INVNEHLKEHLKNQKNFINLEPYFL----FDTIYQGLGIDRIAACYTIEDG- 85
                     : :   : :     :|::|:    |     : ||  |||||  | :
Db         75 PPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPTLGIDRALALWGAGMSW 134

Qy         86 ----VVVDAGSAITIDIISNSIHL-GGFILPGIANYKKIYSHISPRLKSEFNTQVSLDA- 139
                  :|:|||:|:|       :| || ||||:          : :|      || ::|
Db        135 GFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQL-----PQVEMEAI 189

Qy        140 ------FPQKTMDALSYGVFKGIYLLI---KDAAQ-------NKKLYFTGGDGQFLANYF 183
                   |   | :|:  ||   || ||   :|  :       : |:   |||    | ||
Db        190 KSLPPRFALNTTEAIQSGV---IYTLIAGMRDFTEEWLSLFPDGKVAIKGGDRILLLNYL 246

Qy        184 DHAIYDKL---------LIFRGMKKII 201
                |:|  |         ||| ||: |:
Db        247 -QALYPDLAARLIVEPNLIFWGMQTIV 272


RESULT 8
Q98Q93
ID   Q98Q93      PRELIMINARY;      PRT;   224 AA.
AC   Q98Q93;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Hypothetical protein MYPU_4730.
GN   MYPU_4730.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445564; CAC13646.1; -.
DR   MypuList; MYPU_4730; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   224 AA;  25363 MW;  240600CED6AB2B4B CRC64;

  Query Match             11.6%;  Score 126.5;  DB 16;  Length 224;
  Best Local Similarity   21.8%;  Pred. No. 0.0091;
  Matches   41;  Conservative   40;  Mismatches   66;  Indels   41;  Gaps    5;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLE------FKNEQKIFYINVNEHLKEHLKNQKN 54
              :|:| : |:  |              | ||      ||| :  |:|: :: | :  : :
Db         56 ILICSVRNAKEN------------QILEFKLKSIFKNAKIDFFIHKSQSLVKFCQKE-- 100

Qy         55 FINLEPYFLFDTIYQGLGIDRIAACYTI----EDGVVVDAGSAITIDIISNSIHLGGFIL 110
                        :   :|:| :|  | :    :: : :  |:|| |  | ||   |  |
Db        101 -----------LTSEIGLDIVANAYYVLHKSKNAIFISLGTAIVITQIKNSSIEGISIY 148

Qy        111 PGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIK----DAAQN 166
              |||    | : ::  :::| ||         : |:::|  : :|   |:|    :  :
Db        149 PGIYQSFKNFFNVVAKIESNFNIIKIPPILGKTTLESISSSLVRGSVFLLKGYIDEIDKT 208

Qy        167 KKLYFTGG 174
                :: |||
Db        209 SDIFITGG 216


RESULT 9
O51477
ID   O51477      PRELIMINARY;      PRT;   262 AA.
AC   O51477;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein BB0527.
GN   BB0527.
OS   Borrelia burgdorferi (Lyme disease spirochete).
OC   Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX   NCBI_TaxID=139;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35210 / B31;
RX   MEDLINE=98065943; PubMed=9403685;
RA   Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA   Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA   Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA   Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA   van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA   Utterback T., Watthey L., McDonald L., Artiach P., Bowman C.,
RA   Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA   Smith H.O., Venter J.C.;
RT   "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT   burgdorferi.";
RL   Nature 390:580-586(1997).
DR   EMBL; AE001154; AAC66882.1; -.
DR   TIGR; BB0527; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   262 AA;  29822 MW;  DC8CD558CC16B9AF CRC64;

  Query Match             11.1%;  Score 121.5;  DB 16;  Length 262;
  Best Local Similarity   24.9%;  Pred. No. 0.028;
  Matches   62;  Conservative   42;  Mismatches   86;  Indels   59;  Gaps   14;

Qy          2 LLCDIGNSNANF--LDDNKYFTLNIDQFLEFK--------------------NEQKIFYI 39
              |: ||||::  |    ||:     :: |:: |                    |  |:|
Db          9 LIIDIGNTSIAFALFKDNQ-----VNLFIKMKTNLMLRYDEVYSFFEENFDFNVNKVFIS 63

Qy         40 NVNEHLKEHLKNQ-KNFINLEPYFL-FDTIYQ-----------GLGID----RIAAC--Y 80
              :|   | |  ||   :|  ::| |: ||  |             || |     :||   |
Db         64 SVVPILNETFKNVIFSFFKIKPLFIGFDLNYDLTFNPYKSDKFLLGSDVFANLVAAIENY 123

Qy         81 TIEDGVVVDAGSAITIDIISNSIH-LGGFILPG-IANYKKIYSHISPRLKSEFNTQVSLD 138
              : |: :||| |:| ||  :|     ||| |  | : |:  :    :  |  :|      :
Db        124 SFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLD--NAYLIKKFPISTPNN 181
```

```
Qy        139 AFPQKTMDALSYGVF-------KGIYLLIKDAAQNK-KLYFTGGDGQFLANYFD-HAIYD 189
Db        182 LLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITGGNADLILSLIEIEFIFN 241

Qy        190 KLLIFRGMK 198
Db        242 IHLTVEGVR 250
```

RESULT 10
Q9HWC1
ID Q9HWC1 PRELIMINARY; PRT; 248 AA.
AC Q9HWC1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein PA4279.
GN PA4279.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004843; AAG07667.1; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001230; Prenyl_site.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 248 AA; 26758 MW; 609D37480899DA3B CRC64;

Query Match 11.0%; Score 120.5; DB 16; Length 248;
Best Local Similarity 24.1%; Pred. No. 0.032;
Matches 38; Conservative 36; Mismatches 45; Indels 39; Gaps 7;

```
Qy         68 YQGLGIDR----IAACYTIEDG-VVVDAGSAITIDII-SNSIHLGGFILPGIANYKKIYS 121
Db         95 YQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMT------- 147

Qy        122 HISPRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK---- 167
Db        148 ----LMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMA 203

Qy        168 --------KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
Db        204 CELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVGL 241
```

RESULT 11
P74045
ID P74045 PRELIMINARY; PRT; 257 AA.
AC P74045;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein slr0812.
GN SLR0812.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90911; BAA18120.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 257 AA; 27702 MW; 4291E51EE91FB4A2 CRC64;

Query Match 10.7%; Score 116.5; DB 16; Length 257;
Best Local Similarity 30.4%; Pred. No. 0.072;
Matches 52; Conservative 26; Mismatches 66; Indels 27; Gaps 11;

```
Qy         49 LKNQKNFINLEPYFLFDTIYQGLGIDRIAA--CYTIEDGVVVDAGSAITIDIISNSIHL- 105
Db         87 LKNLP-LVNLYPSFGIDRALAGLGTGLTYGFPC------LVVDGGTALTITGFDQDKKLV 139

Qy        106 GGFILPGIA-NYKKIYSHIS--PRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKD 162
Db        140 GGAILPGLGLQLATLGDRLAALPKLEMDQLTELP-DRWALDTPSAIFSGVVYGVLGALQS 198

Qy        163 AAQN-KKLY------FTGGDGQFLANYF-DHA-----IYDKLLIFRGMKKI 200
Db        199 YLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNLIFLGMAAI 249
```

RESULT 12
Q9A6Z1
ID Q9A6Z1 PRELIMINARY; PRT; 261 AA.
AC Q9A6Z1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Transcriptional activator, putative, Baf family.
GN CC1935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AE005867; AAK23910.1; -.
DR TIGR; CC1935; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Complete proteome.
SQ SEQUENCE 261 AA; 27965 MW; C19E60D7B0714EF5 CRC64;

Query Match 10.2%; Score 111; DB 16; Length 261;
Best Local Similarity 26.8%; Pred. No. 0.21;
Matches 52; Conservative 31; Mismatches 64; Indels 47; Gaps 11;

```
Qy          48 HLKN-QKNFINLEPYFLFDTIYQGL----------GIDRI-----AACYTIEDGVVVDAG 91
               :|:|  : : |:||  : :    |:          | ||:     ||       ||:|:|
Db          73 NLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSG 132

Qy          92 SAITIDII-SNSIHLGGFILPGIANYKKIYSHIS----PRL---KSEFNTQVSLDAFPQK 143
               :| | ||: ::    || | ||| |      | :    ||:   :   |  |  |
Db         133 TATTFDIVAADGAFEGGIIAPGI-NLSMQALHEAAAKLPRIAIQRPAGNRIVGTD----- 186

Qy         144 TMDALSYGVFKGIYLLIK--------DAAQNKKLYFTGGDGQFLANYFDHAI-----YDK 190
               |: |:  ||| |   ||:        :  :   :  |||    :|: |: |      :|
Db         187 TVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGG----VASLFEGATDSIDHFDS 242

Qy         191 LLIFRGMKKIIKEN 204
                |  ||: :| : |
Db         243 DLTIRGLLEIYRRN 256


RESULT 13
Q9KGH5
ID   Q9KGH5      PRELIMINARY;      PRT;   254 AA.
AC   Q9KGH5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein BH0086.
GN   BH0086.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001507; BAB03805.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   254 AA;  27907 MW;  0E3B557BA7DAC176 CRC64;

  Query Match             10.1%;  Score 110;  DB 16;  Length 254;
  Best Local Similarity   23.4%;  Pred. No. 0.24;
  Matches   60;  Conservative   47;  Mismatches   93;  Indels   56;  Gaps   14;

Qy           1 MLLCDIGNSN---ANFLDDNKYFTLNI--------DQF-----------LEFKNEQKIF 37
               :|: |:||:|     : |:       :        |::           |:|::   |
Db           2 ILVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGIV 61

Qy          38 YINVNEHLKEHLKNQ-KNFINLEPYFLFDTIYQGL----------GIDRIA-ACYTIE-- 83
                 :|   :   |:   | : :: | :   |  ||          | |||  |   ||
Db          62 ISSVVPPMMFSLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRIVNAVAAIELY 121

Qy          84 --DGVVVDAGSAITIDIIS-NSIHLGGFILPGI-ANYKKIYSHIS--PRLKSEFNTQVSL 137
                   :||| |:| |  :|:    : || | |||  : : :|   |  ||::     ||
Db         122 GYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQV-- 179

Qy         138 DAFPQKTMDALSYGVF-------KGIYLLIKDAAQNK-KLYFTGGDGQFLANYFDHA-IY 188
                 |:|::  |:|        |:   :|  |::: |:  |||  : :    :   :
Db         180 --VGTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVI 237

Qy         189 DKLLIFRGMKKIIKEN 204
               |  |  :|:: | |:|
Db         238 DSFLTLKGLQLIYKKN 253


RESULT 14
```

```
O96204
ID   O96204      PRELIMINARY;      PRT;  3973 AA.
AC   O96204;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 480.3 kDa protein.
GN   PFB0555C.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99021743; PubMed=9804551;
RA   Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA   Koonin E.V., Shallom S., Mason T., Yu K., Fujii C., Pederson J.,
RA   Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pertea M.,
RA   Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA   Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT   "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT   falciparum.";
RL   Science 282:1126-1132(1998).
DR   EMBL; AE001402; AAC71900.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   3973 AA;  480283 MW;  9A7ED9E1196213E5 CRC64;

  Query Match             9.6%;  Score 105;  DB 5;  Length 3973;
  Best Local Similarity   22.1%;  Pred. No. 18;
  Matches   55;  Conservative   42;  Mismatches   88;  Indels   64;  Gaps   11;

Qy           8 NSNANFLDDNKYFTLNIDQFLEFKNEQKI----------------------FYIN-VNE 43
               |:| :  ||: |   |:    : :|  ||                      |:| : |
Db        3066 NNNGDGYDDDGYGEKNVSGIYK-ENNNKINVKGDIYNIDNINVYPLNGKLVSIYLNTLKE 3124

Qy          44 HLKE----HLKNQKNFINLEPYFLFDTIYQGLGIDRIAACYTI--------------EDG 85
                |||    || : ||  |:|  |  : :      :||   | :                :
Db        3125 ILKECYECHLGHMKNNENMEKSFFIEHLLLYFLYNRINTIYELFYNFYFTYLRKKENNND 3184

Qy          86 VVVDAGSAITIDIISNSIH-----LGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAF 140
               :::|  :    ::| | |:     :  |    : : :  | :|:       | |: :
Db        3185 ILLDIVNEHIYNLIGNKIYDQINKINNF----LDDKQYYYFYINTLTFITLNKQICIYII 3240

Qy         141 PQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKI 200
               :| :: | |  |  ||  : |  :|    :   : |:: |  || |:  |::|     |
Db        3241 KKKILNKLIYIPF--IYHSLFDKNKNFTSIYHINNNQYIRNK-DHIIFCSLIVF-----I 3292

Qy         201 IKENPNLLY 209
               ||    |||
Db        3293 IK----LLY 3297


RESULT 15
Q8R663
ID   Q8R663      PRELIMINARY;      PRT;   479 AA.
AC   Q8R663;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Cardiolipin synthetase (EC 2.7.8.-).
GN   FN0923.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G:, Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
```

```
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010600; AAL95119.1; -.
KW   Transferase; Complete proteome.
SQ   SEQUENCE   479 AA;  56509 MW;  5C4CB432BF7CFF70 CRC64;

  Query Match             9.5%;  Score 104;  DB 16;  Length 479;
  Best Local Similarity   24.8%;  Pred. No. 1.6;
  Matches   31;  Conservative   26;  Mismatches   38;  Indels   30;  Gaps    6;

Qy         19 YFTLNIDQFLEFKNEQKIFYINVN--EHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDRI 76
              |  :: |  :   |: ::::   :    ||: : | | |||:| ||:|   : |:| :
Db        104 YLEMSTDNHISSNNDIQVYFTGEDFFPELKKEIANAKKFINME-YFIFQ--FDGIGKE-- 158

Qy         77 AACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSEF-NTQV 135
                   | | ::  |   : :::| : ::|  |                |||  | || |
Db        159 -----IADLLIEKAKEGVEVNLIIDGVNLANF----------------RLKQYFKNTGV 196

Qy        136 SLDAF 140
              :|  |
Db        197 NLHLF 201


Search completed: June 24, 2003, 21:59:37
Job time : 19.0461 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 6.10554 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-21
Perfect score: 1093
Sequence: 1 MLLCDIGNSNANFLDDNKYF.....KLLIFRGMKKIIKENPNLLY 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

otal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	89	8.1	1024	4	US-09-091-117-5	Sequence 5, Appli
2	88.5	8.1	1264	1	US-07-789-915A-6	Sequence 6, Appli
3	88.5	8.1	1264	1	US-08-005-002C-6	Sequence 6, Appli
4	88.5	8.1	1264	1	US-08-487-203A-6	Sequence 6, Appli
5	87	8.0	713	4	US-08-628-434-2	Sequence 2, Appli
6	82.5	7.5	476	4	US-09-316-083-3	Sequence 3, Appli
7	81	7.4	713	4	US-08-628-434-4	Sequence 4, Appli
8	80.5	7.4	1648	2	US-08-662-227-35	Sequence 35, Appl
9	80.5	7.4	1648	4	US-09-017-947-35	Sequence 35, Appl
10	80	7.3	754	4	US-09-521-780-2	Sequence 2, Appli
11	80	7.3	754	4	US-09-521-780-3	Sequence 3, Appli
12	79	7.2	347	4	US-09-134-001C-3298	Sequence 3298, Ap
13	78	7.1	518	3	US-09-043-123-4	Sequence 4, Appli
14	78	7.1	518	4	US-08-973-398-2	Sequence 2, Appli
15	78	7.1	518	5	PCT-US96-05800-7	Sequence 7, Appli
16	77	7.0	1430	3	US-09-008-172-2	Sequence 2, Appli
17	77	7.0	1430	4	US-09-210-361-6	Sequence 6, Appli
18	77	7.0	1430	4	US-09-740-274-6	Sequence 6, Appli
19	76.5	7.0	181	2	US-08-392-625-22	Sequence 22, Appl
20	76.5	7.0	181	2	US-08-466-961A-22	Sequence 22, Appl
21	76.5	7.0	181	2	US-08-645-193B-24	Sequence 24, Appl
22	76.5	7.0	282	4	US-09-134-001C-3902	Sequence 3902, Ap
23	76.5	7.0	332	4	US-09-134-001C-4015	Sequence 4015, Ap
24	76.5	7.0	377	4	US-08-195-705-5	Sequence 5, Appli
25	76.5	7.0	806	1	US-08-270-076A-11	Sequence 11, Appl
26	76	7.0	220	2	US-08-840-683-8	Sequence 8, Appli
27	76	7.0	317	2	US-08-555-722-8	Sequence 8, Appli
28	76	7.0	317	4	US-09-384-301-8	Sequence 8, Appli
29	76	7.0	1451	1	US-08-308-872B-4	Sequence 4, Appli
30	75.5	6.9	1642	2	US-08-662-227-2	Sequence 2, Appli
31	75.5	6.9	1642	4	US-09-017-947-2	Sequence 2, Appli
32	75.5	6.9	1651	1	US-08-447-411-2	Sequence 2, Appli
33	74.5	6.8	695	4	US-09-134-001C-4341	Sequence 4341, Ap
34	74.5	6.8	708	3	US-08-613-009A-22	Sequence 22, Appl
35	74.5	6.8	708	4	US-08-778-570B-28	Sequence 28, Appl
36	74.5	6.8	708	4	US-09-059-584-28	Sequence 28, Appl
37	74	6.8	796	4	US-08-868-699A-2	Sequence 2, Appli
38	74	6.8	796	4	US-09-757-014-2	Sequence 2, Appli
39	74	6.8	806	4	US-09-134-001C-4314	Sequence 4314, Ap
40	73.5	6.7	375	4	US-08-195-705-3	Sequence 3, Appli
41	73.5	6.7	638	4	US-09-228-986-74	Sequence 74, Appl
42	73.5	6.7	896	2	US-08-640-389A-10	Sequence 10, Appl
43	73.5	6.7	896	4	US-08-618-957A-10	Sequence 10, Appl
44	73.5	6.7	898	2	US-08-693-697-36	Sequence 36, Appl
45	73.5	6.7	898	4	US-08-588-189-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-091-117-5
; Sequence 5, Application US/09091117
; Patent No. 6171589
; GENERAL INFORMATION:
; APPLICANT: The University of Melbourne
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and
; TITLE OF INVENTION: Vaccines
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: United States of America
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/091,117
; FILING DATE: 12 JUNE 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU96/00803
; FILING DATE: 13-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PN7127
; FILING DATE: 13-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WINNER, Ellen P.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: +1 303 499 8080
; TELEFAX: +1 303 499 8089
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1024 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Mycoplasma genitalium
US-09-091-117-5

Query Match 8.1%; Score 89; DB 4; Length 1024;
Best Local Similarity 23.9%; Pred. No. 1;
Matches 47; Conservative 30; Mismatches 74; Indels 46; Gaps 9;

Qy 26 QFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP---YFLFD----------TIYQG 70

```
              || | || | :  : |   |    |   |::   |    |: |           |:||
Db        301 QFGEHKNGQPLSLLKVLSGLNNEFKQLFNYV--RPGLGDFVSDLIQSSSQSSNKKTVYQL 358

Qy         71 LGIDRIAACYTIEDGVVVDAGSAI-TIDIISNSIHLGGFILPGIANYKKIYSHISPRLKS 129
              |  ::    : ::|  : :  | :  :||:   |:     :   | ::| :   | | ||:
Db        359 LFENKTTIIHLLQDLNISELNSVLPVVDILFEGINSAESLYQRIQSFKDL---IVPALKA 415

Qy        130 EFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYF---------TGGDGQFLA 180
                :   : ||:|     :|  :  ||  :|       |||  : |          |     ||
Db        416 DKQLK-SLEAIILAVLDNPNTYVFDLVY-------QNKSILFNLLSDFLKNTANTLPFLQ 467

Qy        181 NYFD--------HAIYD 189
                ||        ||:|
Db        468 EQFDIVNHLFANEAIFD 484


RESULT 2
US-07-789-915A-6
; Sequence 6, Application US/07789915A
; Patent No. 5212058
  GENERAL INFORMATION:
    APPLICANT:  Baker, Rohan T.
;   APPLICANT:  Tobias, John W.
;   APPLICANT:  Varshavsky, Alexander
;   TITLE OF INVENTION:  Ubiquitin-Specific Proteases
;   NUMBER OF SEQUENCES:  8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET:  Two Militia Drive
;     CITY:  Lexington
;     STATE:  Massachusetts
;     COUNTRY:  U.S.A.
;     ZIP:  02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/07/789,915A
;     FILING DATE:  19911108
;     CLASSIFICATION:  435
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Brook, David E.
;     REGISTRATION NUMBER:  22,592
;     REFERENCE/DOCKET NUMBER:  MIT-5091AA
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  617-861-6240
;     TELEFAX:  617-861-9540
  INFORMATION FOR SEQ ID NO:  6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1264 amino acids
;     TYPE:  AMINO ACID
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  protein
US-07-789-915A-6

  Query Match             8.1%;  Score 88.5;  DB 1;  Length 1264;
  Best Local Similarity   25.1%;  Pred. No. 1.6;
  Matches   50;  Conservative   16;  Mismatches   72;  Indels   61;  Gaps   10;

Qy          4 CDIGNSN-ANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKN-------- 54
              |:|   | |  |   || | |  |                    || | ||
Db        525 CNISELNEATLLSIYKYETSNKSQV----------------TSNHLTNLKNALRLLAKY 567

Qy         55 --------FINLEPYFLFDTIYQGLGIDR------IAACYTIE----DGVVVDAGSAI-T 95
                     ::: |||      |  | ||       |   |:::     |: :|   |: |
Db        568 TKSDKLKFYVDHEPYRALSQAYDTLSIDESVDEDIIKTAYSVKINDSPGLKLDCDRALYT 627

Qy         96 IDIISNSIHLGGFI---LPGIANYKKIYSHISPRLKSEFNTQVS----LDAFPQKTMDAL 148
              | |   |: |  |:    |  :||       |    : |   |    |  | ||  |
Db        628 IAISKRSLDLFNFLTEECPQFSNY------YGPEKLLQVNENASDETILKIFKQKWFDE- 680

Qy        149 SYGVFKGIYLLIKDAAQNK 167
                 |::    ||  ||  |
Db        681 --NVYEPDQFLILRAALTK 697


RESULT 3
US-08-005-002C-6
; Sequence 6, Application US/08005002C
; Patent No. 5494818
;  GENERAL INFORMATION:
;    APPLICANT:  Baker, Rohan T.
;    APPLICANT:  Tobias, John W.
;    APPLICANT:  Varshavsky, Alexander
;    TITLE OF INVENTION:  Ubiquitin-Specific Proteases
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  Maine
;      COUNTRY:  U.S.A.
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/005,002C
;      FILING DATE:  15-JAN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/789,915
;      FILING DATE:  08-NOV-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  MIT-5091AAZ
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  207-363-0558
;      TELEFAX:  207-363-0528
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1264 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-005-002C-6

  Query Match             8.1%;  Score 88.5;  DB 1;  Length 1264;
  Best Local Similarity   25.1%;  Pred. No. 1.6;
  Matches   50;  Conservative   16;  Mismatches   72;  Indels   61;  Gaps   10;

Qy          4 CDIGNSN-ANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKN-------- 54
              |:|   | |  |   || | |  |                    || | ||
Db        525 CNISELNEATLLSIYKYETSNKSQV----------------TSNHLTNLKNALRLLAKY 567

Qy         55 --------FINLEPYFLFDTIYQGLGIDR------IAACYTIE----DGVVVDAGSAI-T 95
                     ::: |||      |  | ||       |   |:::     |: :|   |: |
Db        568 TKSDKLKFYVDHEPYRALSQAYDTLSIDESVDEDIIKTAYSVKINDSPGLKLDCDRALYT 627

Qy         96 IDIISNSIHLGGFI---LPGIANYKKIYSHISPRLKSEFNTQVS----LDAFPQKTMDAL 148
              | |   |: |  |:    |  :||       |    : |   |    |  | ||  |
Db        628 IAISKRSLDLFNFLTEECPQFSNY------YGPEKLLQVNENASDETILKIFKQKWFDE- 680

Qy        149 SYGVFKGIYLLIKDAAQNK 167
                 |::    ||  ||  |
Db        681 --NVYEPDQFLILRAALTK 697


RESULT 4
US-08-487-203A-6
; Sequence 6, Application US/08487203A
```

```
; Patent No. 5683904
;  GENERAL INFORMATION:
;    APPLICANT:  Baker, Rohan T.
;    APPLICANT:  Tobias, John W.
;    APPLICANT:  Varshavsky, Alexander
;    TITLE OF INVENTION:  Ubiquitin-Specific Proteases
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kevin M. Farrell, P.C.
;      STREET:  P.O. Box 999
;      CITY:  York Harbor
;      STATE:  Maine
;      COUNTRY:  U.S.A.
;      ZIP:  03911
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,203A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/005,002
;      FILING DATE:  15-JAN-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Farrell, Kevin M.
;      REGISTRATION NUMBER:  35,505
;      REFERENCE/DOCKET NUMBER:  MIT-5091A3Z
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  207-363-0558
;      TELEFAX:  207-363-0528
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1264 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-487-203A-6

  Query Match             8.1%;  Score 88.5;  DB 1;  Length 1264;
  Best Local Similarity   25.1%;  Pred. No. 1.6;
  Matches   50;  Conservative   16;  Mismatches   72;  Indels   61;  Gaps   10;

Qy          4 CDIGNSN-ANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKN-------- 54
              |:|   | |  |   || | |  |                    || | ||
Db        525 CNISELNEATLLSIYKYETSNKSQV----------------TSNHLTNLKNALRLLAKY 567

Qy         55 --------FINLEPYFLFDTIYQGLGIDR------IAACYTIE----DGVVVDAGSAI-T 95
                    ::: |||      |  | ||       |   |:::     |: :|   |: |
Db        568 TKSDKLKFYVDHEPYRALSQAYDTLSIDESVDEDIIKTAYSVKINDSPGLKLDCDRALYT 627

Qy         96 IDIISNSIHLGGFI---LPGIANYKKIYSHISPRLKSEFNTQVS----LDAFPQKTMDAL 148
              | |   |: |  |:   |  :||       |   : |   |    |  | ||  |
Db        628 IAISKRSLDLFNFLTEECPQFSNY------YGPEKLLQVNENASDETILKIFKQKWFDE- 680

Qy        149 SYGVFKGIYLLIKDAAQNK 167
                |::    ||  ||  |
Db        681 --NVYEPDQFLILRAALTK 697


RESULT 5
US-08-628-434-2
; Sequence 2, Application US/08628434
; Patent No. 6265567
;  GENERAL INFORMATION:
;    APPLICANT:  Sparling, P. Frederick
;    APPLICANT:  Beucher, Margaret
;    TITLE OF INVENTION:  Isolated FrpB Nucleic Acid Molecule          and Vaccine
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ImClone Systems Incorporated
;      STREET:  180 Varick Street
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10014
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,434
;      FILING DATE:  05-APR-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/418,964
;      FILING DATE:  07-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gallagher, Thomas C.
;      REGISTRATION NUMBER:  37,066
;      REFERENCE/DOCKET NUMBER:  SPA-3-P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 645-1405
;      TELEFAX:  (212) 645-2054
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  713 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-628-434-2

  Query Match             8.0%;  Score 87;  DB 4;  Length 713;
  Best Local Similarity   23.3%;  Pred. No. 1;
  Matches   44;  Conservative   23;  Mismatches   64;  Indels   58;  Gaps    8;

Qy         29 EFKNEQKIFYINVNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDRIAACYTIEDGVVV 88
              | |:::|      ||  |:: :|:|                        ||  | : :
Db        378 EIKDKEKA----TNEEKKKNRENEK---------------------IAKAYRLTNPTKT 411

Qy         89 DAGSAITIDIISNSIH-LGGFILPGIANYK--KIYSH----------------ISPRLK 128
              | |: |       :|| : || | |   |   |: :|                 ||
Db        412 DTGAYI------EAIHEIDGFTLTGGLRYDRFKVKTHDGKTVSSSSLNPSFGVIWQPREH 465

Qy        129 SEFNTQVSLDAFPQKTMDALSYGVFKGIYLLI----KDAAQNKKLYFTGGDGQFLAN--Y 182
                |:  :     :  |||    :|| :      : |:| :: |   || | ||  |
Db        466 WSFSASHNYAGRSPRLYDALQTHGKRGIISIADGTKAERARNTEIGFNYNDGTFAANGSY 525

Qy        183 FDHAIYDKL 191
              |   | | |
Db        526 FRQTIKDAL 534


RESULT 6
US-09-316-083-3
; Sequence 3, Application US/09316083A
; Patent No. 6280942
; GENERAL INFORMATION:
;  APPLICANT: The Institute of Physical and Chemical Research
;  TITLE OF INVENTION: Endonuclease
;  FILE REFERENCE: PH-651
;  CURRENT APPLICATION NUMBER: US/09/316,083A
;  CURRENT FILING DATE: 1999-05-20
;  EARLIER APPLICATION NUMBER: JP98/141861
;  EARLIER FILING DATE: 1998-05-22
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;  LENGTH: 476
;  TYPE: PRT
;  ORGANISM: Saccharomyces cerevisiae
US-09-316-083-3
```

```
  Query Match             7.5%;  Score 82.5;  DB 4;  Length 476;
  Best Local Similarity   20.8%;  Pred. No. 1.7;
  Matches   50;  Conservative   36;  Mismatches   79;  Indels   75;  Gaps   11;

Qy         12 NFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP-YFLFDTI-YQ 69
              | :::  | | ||:      |    ||: ::::|       : :|: | ::|: | |:
Db        131 NIMNNLNYKTSNIE-----TNLSNNFYL-MDKYLINKYMKYLDMLNMIPNNYMFNNINYK 184

Qy         70 GLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKS 129
              |         :   |:|  :    | :|  |   |:| ||      | :| :  | :
Db        185 G----------KLNIKTVLDLNNNEFYDYLSGLIEGDGYIGPGGIT---ITNHANDVLNT 231

Qy        130 EF-NTQVSLDAFPQKTMDALSYG----------------VFKGIY------------ 157
              | | ::      :| || |                    :|  ||
Db        232 IFINKRIKNSILVEKWMDTLKDNPYFVNAFSINIKTNLAKEKIFTNIYNKLYSDYKINQI 291

Qy        158 --------------LLIKDAAQNKKLY----FTGGDGQFLANYFDHAIYDKLLIFRGMK 198
                           | ||:    |  |    ||  || ||:: ::       |:|: |:
Db        292 NNHIPYYNYLKINNKLPIKNIMDIKNNYWLAGFTAADGSFLSSMYNP---KDTLLFKNMR 348


RESULT 7
US-08-628-434-4
; Sequence 4, Application US/08628434
; Patent No. 6265567
;  GENERAL INFORMATION:
;    APPLICANT:  Sparling, P. Frederick
;    APPLICANT:  Beucher, Margaret
;    TITLE OF INVENTION:  Isolated FrpB Nucleic Acid Molecule           and Vaccine
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ImClone Systems Incorporated
;      STREET:  180 Varick Street
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10014
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,434
;      FILING DATE:  05-APR-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/418,964
;      FILING DATE:  07-APR-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gallagher, Thomas C.
;      REGISTRATION NUMBER:  37,066
;      REFERENCE/DOCKET NUMBER:  SPA-3-P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 645-1405
;      TELEFAX:  (212) 645-2054
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  713 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-628-434-4

  Query Match             7.4%;  Score 81;  DB 4;  Length 713;
  Best Local Similarity   23.8%;  Pred. No. 4.3;
  Matches   45;  Conservative   24;  Mismatches   64;  Indels   56;  Gaps    9;

Qy         29 EFKNEQKIFYINVNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDRIAACYTIEDGVVV 88
              :|| | |    :  |  |:  ||::|                  ::||  | : :
Db        376 QFKIEDK---KDATEEDKK--KNREN------------------EKIAKAYRLTNPTKT 411

Qy         89 DAGSAITIDIISNSIH-LGGFILPGIANYK--KIYSH----------------ISPRLK 128
                                                                  | |: |
Db        412 DTGAYI------EAIHEIDGFTLTGGLRYDRFKVKTHDGKTVSSSSLNPSFGVIWQPREH 465

Qy        129 SEFNTQVSLDAFPQKTMDALSYGVFKGIYLLI----KDAAQNKKLYFTGGDGQFLAN--Y 182
                |:   :  :   :  |||     :|| :      : |:| :: |   || | ||  |
Db        466 WSFSASHNYASRSPRLYDALQTHGKRGIISIADGTKAERARNTEIGFNYNDGTFAANGSY 525

Qy        183 FDHAIYDKL 191
              |   | | |
Db        526 FRQTIKDAL 534


RESULT 8
US-08-662-227-35
; Sequence 35, Application US/08662227
; Patent No. 5922320
;  GENERAL INFORMATION:
;    APPLICANT:  VOGEL, CARL-WILHELM
;    APPLICANT:  BREDEHORST, REINHORST
;    APPLICANT:  KOCK, MICHAEL
;    APPLICANT:  FRITZINGER, DAVID
;    TITLE OF INVENTION:  RECOMBINANT PROCVF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/662,227
;      FILING DATE:  14-JUN-1996
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-0107-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1648 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-662-227-35

  Query Match             7.4%;  Score 80.5;  DB 2;  Length 1648;
  Best Local Similarity   23.2%;  Pred. No. 16;
  Matches   45;  Conservative   25;  Mismatches   69;  Indels   55;  Gaps    8;

Qy         10 NANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEPYFLFDTI-- 67
              ||| |   ||||  |       |: ||| :      ::  :: :| : :  :   | | .
Db        469 NANSLKQIKYFTYLI------LNKGKIFKVG-----RQPRRDGQNLVTMNLHITPDLIPS 517

Qy         68 -----YQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSH 122
                  | :| : | |     | | ||        ::    :|    :|| |
Db        518 PRFVAYYQVGNNEIVA-----DSVWVDVKDTCMGTLVVKGDNL--IQMPGAA-------- 562

Qy        123 ISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLL----------IKDAAQNKKLYFT 172
              :  :|: :|  :| | |             | | :|:|          | |  :      |
Db        563 MKIKLEGDFGARVGLVA-----------VDKAVYVLNDKYKISQAKIWDTIEKSDFGCT 610

Qy        173 GGDGQFLANYFDHA 186
```

```
                    | ||      |: |
Db        611 AGSGQNNLGVFEDA 624


RESULT 9
US-09-017-947-35
; Sequence 35, Application US/09017947
; Patent No. 6303754
;  GENERAL INFORMATION:
;    APPLICANT:  VOGEL, CARL-WILHELM
;    APPLICANT:  BREDEHORST, REINHORST
;    APPLICANT:  KOCK, MICHAEL
;    APPLICANT:  FRITZINGER, DAVID
;    TITLE OF INVENTION:  RECOMBINANT PROCVF
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. JEFFERSON DAVIS HIGHWAY
;      CITY:  ARLINGTON
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/017,947
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/662,227
;      FILING DATE:  14-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F.
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  1126-0107-0X
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  703-413-3000
;      TELEFAX:  703-413-2220
;  INFORMATION FOR SEQ ID NO:  35:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1648 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-017-947-35

  Query Match             7.4%;  Score 80.5;  DB 4;  Length 1648;
  Best Local Similarity   23.2%;  Pred. No. 16;
  Matches   45;  Conservative   25;  Mismatches   69;  Indels   55;  Gaps    8;

Qy         10 NANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEPYFLFDTI-- 67
              ||| |   ||||  |       |: ||| :      ::  :: :| : :  :   | |
Db        469 NANSLKQIKYFTYLI------LNKGKIFKVG-----RQPRRDGQNLVTMNLHITPDLIPS 517

Qy         68 -----YQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSH 122
                  |  :| : | |     | | ||       ::    :|    :|| |
Db        518 PRFVAYYQVGNNEIVA-----DSVWVDVKDTCMGTLVVKGDNL--IQMPGAA-------- 562

Qy        123 ISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLL----------IKDAAQNKKLYFT 172
              :  :|: :|  :| | |            | | :|:|         | |  :      |
Db        563 MKIKLEGDFGARVGLVA------------VDKAVYVLNDKYKISQAKIWDTIEKSDFGCT 610

Qy        173 GGDGQFLANYFDHA 186
                | ||     |: |
Db        611 AGSGQNNLGVFEDA 624


RESULT 10
```

```
US-09-521-780-2
; Sequence 2, Application US/09521780
; Patent No. 6376658
; GENERAL INFORMATION:
;  APPLICANT: Park, Chong-Hun; Samyang Genex Corporation
;  TITLE OF INVENTION: Encapsulation-relating proteins and genes coding the same
;  FILE REFERENCE: ERP
;  CURRENT APPLICATION NUMBER: US/09/521,780
;  CURRENT FILING DATE: 2000-03-09
;  EARLIER APPLICATION NUMBER: KR 1999-27933
;  EARLIER FILING DATE: 1999-07-07
;  EARLIER APPLICATION NUMBER: KR 1999-31172
;  EARLIER FILING DATE: 1999-07-26
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: KOPATIN 1.5
; SEQ ID NO 2
;   LENGTH: 754
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
US-09-521-780-2

  Query Match             7.3%;  Score 80;  DB 4;  Length 754;
  Best Local Similarity   23.0%;  Pred. No. 6;
  Matches   64;  Conservative   25;  Mismatches   81;  Indels  108;  Gaps   15;

Qy          5 DIGNSNANFLDDNKY--FTLNIDQF-LEFKNEQKIFYINVNEHLK----EHLKN------ 51
              |||  |: :   | |  | :: ::| |:  |  ::||    : |     | | |
Db        218 DIG-VNSFYYYYNLYYPFWMSGEEFNLKADNRGELFYYMYQQILARYYLERLSNGFGEIK 276

Qy         52 -------------------------QKNFINLEPYFLFDTIYQGLGIDRIAACYT---- 81
                                       : ||  || ||   |  |  | :|    ||
Db        277 TFNWEVPLETGYYPSLVYPNGLQFPTRPNFAKLEDYFY--TYGQKYGDNRYVYSYTYVKD 334

Qy         82 ----IEDGVVVDAGSAIT------------IDIISNSI---------------HLGGFI 109
                  | |  |:| |   |             :|:: | |                |:
Db        335 YERRIRD--VIDLGYVYTKSGQKVELFSSEGLDVLGNLIECSMDSPYYKFYGAYHMFARH 392

Qy        110 LPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKG-IYLLIKDAAQNKK 168
              | | |      |    |     | | :   ||          | :||  ||   :  ||: |
Db        393 LLGYAYQPYTYHKSQPSALEHFETSMRDPAF---------YQLFKKIIYHFFRFKAQHYK 443

Qy        169 LY---------FTGGDGQF--LANYFD-------HAIY 188
                |          |  : :|  |  |||       :|:|
Db        444 AYTMKDLLYEGVTVKNVEFDRLVTYFDKYYADLTNAVY 481


RESULT 11
US-09-521-780-3
; Sequence 3, Application US/09521780
; Patent No. 6376658
; GENERAL INFORMATION:
;  APPLICANT: Park, Chong-Hun; Samyang Genex Corporation
;  TITLE OF INVENTION: Encapsulation-relating proteins and genes coding the same
;  FILE REFERENCE: ERP
;  CURRENT APPLICATION NUMBER: US/09/521,780
;  CURRENT FILING DATE: 2000-03-09
;  EARLIER APPLICATION NUMBER: KR 1999-27933
;  EARLIER FILING DATE: 1999-07-07
;  EARLIER APPLICATION NUMBER: KR 1999-31172
;  EARLIER FILING DATE: 1999-07-26
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: KOPATIN 1.5
; SEQ ID NO 3
;   LENGTH: 754
;   TYPE: PRT
;   ORGANISM: Tenebrio molitor
;   FEATURE:
;   NAME/KEY: CHAIN
;   LOCATION: (18)..(754)
US-09-521-780-3

  Query Match             7.3%;  Score 80;  DB 4;  Length 754;
  Best Local Similarity   23.0%;  Pred. No. 6;
```

```
  Matches   64;  Conservative   25;  Mismatches   81;  Indels  108;  Gaps   15;

Qy          5 DIGNSNANFLDDNKY--FTLNIDQF-LEFKNEQKIFYINVNEHLK----EHLKN------ 51
              |||  |: :   | |  | :: ::| |:  |  ::||    : |     | | |
Db        218 DIG-VNSFYYYYNLYYPFWMSGEEFNLKADNRGELFYYMYQQILARYYLERLSNGFGEIK 276

Qy         52 -------------------------QKNFINLEPYFLFDTIYQGLGIDRIAACYT---- 81
                                       : ||  || ||   |  |  | :|    ||
Db        277 TFNWEVPLETGYYPSLVYPNGLQFPTRPNFAKLEDYFY--TYGQKYGDNRYVYSYTYVKD 334

Qy         82 ----IEDGVVVDAGSAIT------------IDIISNSI---------------HLGGFI 109
                  | |  |:| |   |              :|:: | |               |:
Db        335 YERRIRD--VIDLGYVYTKSGQKVELFSSEGLDVLGNLIECSMDSPYYKFYGAYHMFARH 392

Qy        110 LPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKG-IYLLIKDAAQNKK 168
              | | |     |    |     | | :   ||         | :||  ||   :  ||: |
Db        393 LLGYAYQPYTYHKSQPSALEHFETSMRDPAF---------YQLFKKIIYHFFRFKAQHYK 443

Qy        169 LY---------FTGGDGQF--LANYFD-------HAIY 188
               |          |  : :|  |  |||       :|:|
Db        444 AYTMKDLLYEGVTVKNVEFDRLVTYFDKYYADLTNAVY 481


RESULT 12
US-09-134-001C-3298
; Sequence 3298, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3298
;   LENGTH: 347
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3298

  Query Match             7.2%;  Score 79;  DB 4;  Length 347;
  Best Local Similarity   24.2%;  Pred. No. 2.5;
  Matches   39;  Conservative   27;  Mismatches   59;  Indels   36;  Gaps    9;

Qy          7 GNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEPYF---- 62
              |:   |:| :     |  |  :::| :       |  | ||| : |      :: ||
Db        191 GDKKVNYLKNE----LRFDAGIDYKKD------NFPEALKEAVPN-----GIDVYFENVG 235

Qy         63 --LFDTIYQGLGID-RIAACYTIE--DGVVVDAGSAITIDIISNSIHLGGFILPGIANYK 117
                : | ::: |    ||  |  |   :    | |  |   :| |   : |||   :| :
Db        236 GYIGDEVFKHLNTHARIPVCGAISSYNHPEKDIGPRIQQTLIKNQAMMRGFI---VAEFA 292

Qy        118 KIYSHISPRL-----KSEFNTQVSL----DAFPQKTMDALS 149
               :   | :|     :::  ||||:     |  ||    : |:
Db        293 DGFKEASKQLAQWVQENKIKTQVSVEDGFDKVPQAFRNLLT 333


RESULT 13
US-09-043-123-4
; Sequence 4, Application US/09043123A
; Patent No. 6096521
; GENERAL INFORMATION:
;  APPLICANT: HAAS, Rainer
;  APPLICANT: ODENBREIT, Stefan
;  APPLICANT: MEYER, Thomas F.
;  APPLICANT: BLUM, Andre
;  APPLICANT: CORTHESY-THEULAZ, Irene
;  TITLE OF INVENTION: NEW ADHESIN FROM HELICOBACTER PYLORI
;  FILE REFERENCE: 05648004
;  CURRENT APPLICATION NUMBER: US/09/043,123A
;  CURRENT FILING DATE: 1998-06-26
;  EARLIER APPLICATION NUMBER: DE/195 35 321.8
;  EARLIER FILING DATE: 1995-09-22
;  NUMBER OF SEQ ID NOS: 5
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 518
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-043-123-4

  Query Match             7.1%;  Score 78;  DB 3;  Length 518;
  Best Local Similarity   18.2%;  Pred. No. 5.7;
  Matches   56;  Conservative   50;  Mismatches   89;  Indels  112;  Gaps   14;

Qy          2 LLCDIGN-----SNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLK---------- 46
              :|  ||:     :| ||          |: | | : |   :|:  :|: |:
Db        148 ILSTIGSQTNYGTNTNF--------PNMQQQLTYLNAGNVFFNAMNKALENKNGTSSASG 199

Qy         47 ----------------EHLKNQKNFIN-----------LEPY--------------F 62
                             ::|: |:| :|            || :
Db        200 TSGATGSDGQTYSTQAIQYLQGQQNILNNAANLLKQDELLLEAFNSAVAANIGNKEFNSA 259

Qy         63 LFDTIYQGLGIDRIAACY------TIEDGVVVDAG----------------------SAI 94
               |  : ||: ||:  | |       ||    |: ||                      : :
Db        260 AFTGLVQGI-IDQSQAVYNELTKNTISGSAVISAGINSNQANAVQGRASQLPNALYNAQV 318

Qy         95 TIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAF--PQKTMDALSYGV 152
              |:| | |:::     :| :  ::   |  |  : : | |::    |   :: :    ||
Db        319 TLDKI-NALNNQVRSMPYLPQFRAGNSR-STNILNGFYTKIGYKQFFGKKRNIGLRYYGF 376

Qy        153 F--KGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIYDKLL---IFRGM-------KKI 200
              |   |  :  :   |  ||  |     | | |  :  :: :    | |:       :
Db        377 FSYNGASVGFRSTQNNVGLYTYGVGTDVLYNIFSRSYQNRSVDMGFFSGIQLAGETFQST 436

Qy        201 IKENPNL 207
              ::::||:
Db        437 LRDDPNV 443


RESULT 14
US-08-973-398-2
; Sequence 2, Application US/08973398
; Patent No. 6468739
;  GENERAL INFORMATION:
;    APPLICANT:  HAAS, Ranier
;    APPLICANT:  ODENBREIT, Stefan
;    APPLICANT:  MEYER, Thomas F.
;    TITLE OF INVENTION:  PROCESS FOR IDENTIFYING SECRETORY GENES FROM
;    TITLE OF INVENTION:  HELICOBACTER PYLORI
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nikaido, Marmelstein, Murray & Oram LLP
;      STREET:  655 Fifteenth Street N.W. Suite 330
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-5701
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/973,398
;      FILING DATE:  11-MAR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO PCT/EP96/02544
;      FILING DATE:  12-JUN-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE 19521312.2
```

```
;      FILING DATE:  12-JUN-1995
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  518 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-973-398-2

  Query Match             7.1%;  Score 78;  DB 4;  Length 518;
  Best Local Similarity   18.2%;  Pred. No. 5.7;
  Matches   56;  Conservative   50;  Mismatches   89;  Indels  112;  Gaps   14;

Qy          2 LLCDIGN-----SNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLK---------- 46
              :|  ||:     :| ||         |: | | : |   :|:  :|: |:
Db        148 ILSTIGSQTNYGTNTNF--------PNMQQQLTYLNAGNVFFNAMNKALENKNGTSSASG 199

Qy         47 ----------------EHLKNQKNFIN-----------LEPY--------------F 62
                            ::|: |:| :|            || :
 b        200 TSGATGSDGQTYSTQAIQYLQGQQNILNNAANLLKQDELLLEAFNSAVAANIGNKEFNSA 259

Qy         63 LFDTIYQGLGIDRIAACY------TIEDGVVVDAG----------------------SAI 94
               |  : ||: ||:  | |      ||    |: ||                      : :
Db        260 AFTGLVQGI-IDQSQAVYNELTKNTISGSAVISAGINSNQANAVQGRASQLPNALYNAQV 318

Qy         95 TIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAF--PQKTMDALSYGV 152
              |:| | |:::     :| :  ::   |  |  : : | |::    |   :: :    ||
Db        319 TLDKI-NALNNQVRSMPYLPQFRAGNSR-STNILNGFYTKIGYKQFFGKKRNIGLRYYGF 376

Qy        153 F--KGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIYDKLL---IFRGM-------KKI 200
              |   |  :  :   |  ||  |     | | |  :  :: :    | |:       :
Db        377 FSYNGASVGFRSTQNNVGLYTYGVGTDVLYNIFSRSYQNRSVDMGFFSGIQLAGETFQST 436

Qy        201 IKENPNL 207
              ::::||:
Db        437 LRDDPNV 443


RESULT 15
PCT-US96-05800-7
; Sequence 7, Application PC/TUS9605800
;  GENERAL INFORMATION:
;    APPLICANT:  OraVax, Inc.
;    TITLE OF INVENTION:  MULTIMERIC, RECOMBINANT UREASE VACCINE
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Fish & Richardson P.C.
;      STREET:  225 Franklin Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110-2804
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/05800
;      FILING DATE:  23-APR-1996
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/431,041
;      FILING DATE:  28-APR-1995
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/568,122
;      FILING DATE:  06-DEC-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Paul T.
;      REGISTRATION NUMBER:  30,162
;      REFERENCE/DOCKET NUMBER:  06132/020001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 542-5070
;      TELEFAX:  (617) 542-8906
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  518 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US96-05800-7

  Query Match             7.1%;  Score 78;  DB 5;  Length 518;
  Best Local Similarity   18.2%;  Pred. No. 5.7;
  Matches   56;  Conservative   50;  Mismatches   89;  Indels  112;  Gaps   14;

Qy          2 LLCDIGN-----SNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLK---------- 46
              :|  ||:     :| ||         |: | | : |   :|:  :|: |:
Db        148 ILSTIGSQTNYGTNTNF--------PNMQQQLTYLNAGNVFFNAMNKALENKNGTSSASG 199

Qy         47 ----------------EHLKNQKNFIN-----------LEPY--------------F 62
                            ::|: |:| :|            || :
Db        200 TSGATGSDGQTYSTQAIQYLQGQQNILNNAANLLKQDELLLEAFNSAVAANIGNKEFNSA 259

Qy         63 LFDTIYQGLGIDRIAACY------TIEDGVVVDAG----------------------SAI 94
               |  : ||: ||:  | |      ||    |: ||                      : :
Db        260 AFTGLVQGI-IDQSQAVYNELTKNTISGSAVISAGINSNQANAVQGRASQLPNALYNAQV 318

Qy         95 TIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAF--PQKTMDALSYGV 152
              |:| | |:::     :| :  ::   |  |  : : | |::    |   :: :    ||
Db        319 TLDKI-NALNNQVRSMPYLPQFRAGNSR-STNILNGFYTKIGYKQFFGKKRNIGLRYYGF 376

Qy        153 F--KGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIYDKLL---IFRGM-------KKI 200
              |   |  :  :   |  ||  |     | | |  :  :: :    | |:       :
Db        377 FSYNGASVGFRSTQNNVGLYTYGVGTDVLYNIFSRSYQNRSVDMGFFSGIQLAGETFQST 436

Qy        201 IKENPNL 207
              ::::||:
Db        437 LRDDPNV 443


Search completed: June 24, 2003, 21:49:34
Job time : 8.10554 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 9.52888 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-21
Perfect score: 1093
Sequence: 1 MLLCDIGNSNANFLDDNKYF.....KLLIFRGMKKIIKENPNLLY 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1093	100.0	209	9	US-09-813-453A-21	Sequence 21, Appl
2	391.5	35.8	223	9	US-09-895-913A-74	Sequence 74, Appl
3	391.5	35.8	223	9	US-09-813-453A-14	Sequence 14, Appl
4	391.5	35.8	223	9	US-09-813-453A-67	Sequence 67, Appl
5	177.5	16.2	229	9	US-09-813-453A-12	Sequence 12, Appl
6	139.5	12.8	246	9	US-09-813-453A-9	Sequence 9, Appli
7	126.5	11.6	244	9	US-09-813-453A-41	Sequence 41, Appl
8	124	11.3	241	9	US-09-813-453A-63	Sequence 63, Appl
9	121.5	11.1	262	9	US-09-813-453A-11	Sequence 11, Appl
10	120.5	11.0	248	9	US-09-813-453A-20	Sequence 20, Appl
11	116.5	10.7	257	9	US-09-813-453A-13	Sequence 13, Appl
12	114.5	10.5	249	9	US-09-813-453A-61	Sequence 61, Appl
13	113.5	10.4	249	9	US-09-813-453A-70	Sequence 70, Appl
14	113	10.3	262	9	US-09-813-453A-45	Sequence 45, Appl
15	111	10.2	260	9	US-09-813-453A-51	Sequence 51, Appl
16	110	10.1	254	9	US-09-813-453A-47	Sequence 47, Appl
17	109	10.0	258	9	US-09-813-453A-2	Sequence 2, Appli
18	108	9.9	257	9	US-09-813-453A-53	Sequence 53, Appl
19	96	8.8	258	9	US-09-813-453A-6	Sequence 6, Appli
20	91.5	8.4	272	9	US-09-712-363-276	Sequence 276, App
21	91.5	8.4	272	9	US-09-813-453A-5	Sequence 5, Appli
22	89.5	8.2	242	9	US-09-813-453A-65	Sequence 65, Appl
23	89.5	8.2	256	9	US-09-769-787-93	Sequence 93, Appl
24	89.5	8.2	256	10	US-09-815-242-13326	Sequence 13326, A
25	88.5	8.1	592	9	US-09-813-453A-43	Sequence 43, Appl
26	87	8.0	713	9	US-09-801-451A-2	Sequence 2, Appli
27	86.5	7.9	219	9	US-09-813-453A-57	Sequence 57, Appl
28	85	7.8	212	9	US-09-813-453A-59	Sequence 59, Appl
29	85	7.8	592	9	US-09-813-453A-22	Sequence 22, Appl
30	84.5	7.7	272	10	US-09-815-242-10490	Sequence 10490, A
31	82.5	7.5	256	9	US-09-813-453A-55	Sequence 55, Appl
32	82.5	7.5	460	9	US-09-813-453A-39	Sequence 39, Appl
33	82.5	7.5	476	10	US-09-774-414-3	Sequence 3, Appli
34	81	7.4	713	9	US-09-801-451A-4	Sequence 4, Appli
35	80.5	7.4	1648	10	US-09-925-442-35	Sequence 35, Appl
36	78.5	7.2	376	9	US-09-978-756-2	Sequence 2, Appli
37	78.5	7.2	376	9	US-10-087-464-34	Sequence 34, Appl
38	78.5	7.2	378	9	US-10-087-464-11	Sequence 11, Appl
39	78.5	7.2	1639	9	US-10-087-464-10	Sequence 10, Appl
40	78	7.1	258	9	US-09-813-453A-49	Sequence 49, Appl
41	78	7.1	518	9	US-09-976-297-4	Sequence 4, Appli
42	77.5	7.1	456	10	US-09-997-664-92	Sequence 92, Appl
43	77	7.0	545	9	US-09-995-749A-10	Sequence 10, Appl
44	77	7.0	1430	10	US-09-740-274-6	Sequence 6, Appli
45	76.5	7.0	377	9	US-10-062-809-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-09-813-453A-21
; Sequence 21, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
;   LENGTH: 209
;   TYPE: PRT
;   ORGANISM: Campylobacter jejuni
US-09-813-453A-21

  Query Match            100.0%;  Score 1093;  DB 9;  Length 209;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-97;
  Matches  209;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60

Qy         61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120

Qy        121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180

Qy        181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
              |||||||||||||||||||||||||||||
```

```
Db        181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209


RESULT 2
US-09-895-913A-74
; Sequence 74, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
;  APPLICANT: Kleanthous, Harold
;  APPLICANT:  Al-Garawi, Amal
;  APPLICANT:  Miller, Charles
;  APPLICANT:  Tomb, Jean Francois
;  APPLICANT:  Oomen, Raymond P.
;  TITLE OF INVENTION: Identification of Polynucleotides
;  TITLE OF INVENTION:  Encoding No. US20020160456A1el Helicobacter Polypeptides in the
;  TITLE OF INVENTION:  Genome
;  FILE REFERENCE: 06132/043002
;  CURRENT APPLICATION NUMBER: US/09/895,913A
;  CURRENT FILING DATE:  2001-06-29
;  PRIOR APPLICATION NUMBER: US 08/881,227
   PRIOR FILING DATE: 1997-06-24
   NUMBER OF SEQ ID NOS: 368
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-895-913A-74

  Query Match           35.8%;  Score 391.5;  DB 9;  Length 223;
  Best Local Similarity  39.2%;  Pred. No. 5.4e-30;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::|||||:  :|  : : |:   :       :::||||:|||  :: | |   | |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   : |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 3
US-09-813-453A-14
; Sequence 14, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-813-453A-14

  Query Match           35.8%;  Score 391.5;  DB 9;  Length 223;
  Best Local Similarity  39.2%;  Pred. No. 5.4e-30;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::|||||:  :|  : : |:   :       :::||||:|||  :: | |   | |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   : |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 4
US-09-813-453A-67
; Sequence 67, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
;   LENGTH: 223
;   TYPE: PRT
;   ORGANISM: Helicobacter pylori
US-09-813-453A-67

  Query Match           35.8%;  Score 391.5;  DB 9;  Length 223;
  Best Local Similarity  39.2%;  Pred. No. 5.4e-30;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::|||||:  :|  : : |:   :       :::||||:|||  :: | |   | |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   : |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 5
US-09-813-453A-12
; Sequence 12, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
```

```
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Aquifex aeolicus
US-09-813-453A-12

  Query Match             16.2%;  Score 177.5;  DB 9;  Length 229;
  Best Local Similarity   25.8%;  Pred. No. 2.1e-09;
  Matches   62;  Conservative   44;  Mismatches   83;  Indels   51;  Gaps   10;

Qy          2 LLCDIGNSNANF-LDDNKYFTLNIDQFLEFKNEQ---------KIFYINVNEHLKEHLK- 50
              |  |:|||: :  | : |      :  ||:  :|:         |   |:| :   | ::
Db          4 LTVDVGNSSVDIALWEGK----KVKDFLKLSHEEFLKEEFPKLKALGISVKQSFSEKVRG 59

Qy         51 --------NQKNFINLEPYFLFDTIYQGLGIDRIAACYTI-----EDGVVVDAGSAITID 97
                     ::||      |   :|    || ||:|  |:      :: ||: ||:|: ||
Db         60 KIPKIKFLKKENFPIQVDYKTPET----LGTDRVALAYSAKKFYGKNVVVISAGTALVID 115

Qy         98 IISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIY 157
              ::      ||||  |:    || | ::  :   |  :| :    : | : :  | ::
Db        116 LVLEGKFKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEI-FLGRSTRECVLGGAYREST 174

Qy        158 LLIKDA--------AQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
               ||            :  |:  |||:|:: : :    |||  ||: |||:       ||||
Db        175 EFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGMR-------NLLY 224


RESULT 6
US-09-813-453A-9
; Sequence 9, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Thermotoga maritima
US-09-813-453A-9

  Query Match             12.8%;  Score 139.5;  DB 9;  Length 246;
  Best Local Similarity   22.7%;  Pred. No. 1e-05;
  Matches   58;  Conservative   38;  Mismatches   82;  Indels   77;  Gaps   10;

Qy          2 LLCDIGNSNANF--LDDNKYF---------------------TLNIDQFLEFKN------ 32
              || |:||::: |   :| | |                     |  |   | |
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         33 -----------EQKIFYIN------VNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDR 75
                        || |:|:       |  :| ::||                  :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNVKNPSE---------------VGADR 106

Qy         76 IAACYTI-----EDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSE 130
                 :|          ::|:::| |:| |:|:: |  : || ||||   :  ::|      |
Db        107 VANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGF--FMMVHSLFRGTAKLP 164

Qy        131 FNTQVSLDAFPQK-TMDALSYGVFKG-IYLL------IKDAAQNKKLYFTGGDGQFLANY 182
                      |    | | : :  ||  | :| |       ||:   :  :  |||  : : :
Db        165 LVEVKPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDM 224

Qy        183 FDHAIYDKLLIFRGM 197
                  | |:|: |  :|:
Db        225 IKHEIFDEDLTIKGV 239


RESULT 7
US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 244
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

  Query Match             11.6%;  Score 126.5;  DB 9;  Length 244;
  Best Local Similarity   28.0%;  Pred. No. 0.00018;
  Matches   45;  Conservative   29;  Mismatches   50;  Indels   37;  Gaps    8;

Qy         71 LGIDRIAA------CYTIEDGVVVDAGSAITIDIIS-NSIHLGGFILPGI-ANYKKIY-- 120
              || ||:||       |   : :|:|||:||| : :|   |:||| | ||:   :| ::
Db         93 LGADRLAAVVGAHSLYPNTELLVIDAGTAITYERVSAEGIYLGGNISPGLHLRFKALHLF 152

Qy        121 ---------SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGI------YLLIKDAAQ 165
                      | |||::          :   | :|:: ||  |:      |:    | :
Db        153 TGRLPLIDPSGISPKIAE----------YGSSTEEAITAGVIHGLAGEIDRYIDDLHAKE 202

Qy        166 NKK-LYFTGGDGQFLANYFDHAI-YDKLLIFRGMKKIIKEN 204
               :  :  ||||  :||       |     |:  |: :|:: |
Db        203 GRSAVILTGGDANYLARIIRSGILIHPDLVLLGLNRILEYN 243


RESULT 8
US-09-813-453A-63
; Sequence 63, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
```

```
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
;   LENGTH: 241
;   TYPE: PRT
;   ORGANISM: Thiobacillus ferrooxidans
US-09-813-453A-63

  Query Match             11.3%;  Score 124;  DB 9;  Length 241;
  Best Local Similarity   28.4%;  Pred. No. 0.00031;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   14;  Gaps    4;

Qy         69 QGLGIDR----IAAC--YTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSH 122
              : || ||    :||   |  :| :|:| |:|||||:::     || ||||||   :   |
Db         95 ESLGFDRRCCLLAAAMDYPGQDSIVIDMGTAITIDLLAGGHFRGGRILPGIAMSLR-GLH 153

Qy        123 ISPRLKSEFNTQVSLDAFPQKTMDALSYGV-------FKGIYLLIKDAAQNKKLYFTGGD 175
                |  |       :     | :|:  ||        :|     :  :   ::  ||||
Db        154 EGTALLPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGD 213

Qy        176 GQFLANYFDHAIYDKLLIFRG 196
              :     ::|   |: ||
Db        214 AERWQPGIAGSLYQPHLLLRG 234


RESULT 9
US-09-813-453A-11
; Sequence 11, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Borrelia burgdorferi
US-09-813-453A-11

  Query Match             11.1%;  Score 121.5;  DB 9;  Length 262;
  Best Local Similarity   24.9%;  Pred. No. 0.0006;
  Matches   62;  Conservative   42;  Mismatches   86;  Indels   59;  Gaps   14;

Qy          2 LLCDIGNSNANF--LDDNKYFTLNIDQFLEFK--------------------NEQKIFYI 39
              |: ||||::  |    ||:     :: |:: |                   |  |:|
Db          9 LIIDIGNTSIAFALFKDNQ-----VNLFIKMKTNLMLRYDEVYSFFEENFDFNVNKVFIS 63

Qy         40 NVNEHLKEHLKNQ-KNFINLEPYFL-FDTIYQ-----------GLGID----RIAAC--Y 80
              :|   | |  ||   :|  ::| |: ||  |              || |     :||   |
Db         64 SVVPILNETFKNVIFSFFKIKPLFIGFDLNYDLTFNPYKSDKFLLGSDVFANLVAAIENY 123

Qy         81 TIEDGVVVDAGSAITIDIISNSIH-LGGFILPG-IANYKKIYSHISPRLKSEFNTQVSLD 138
              : |: :||| |:| ||  :|     ||| |  | : |:  :   :  |  :|      :
Db        124 SFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLD--NAYLIKKFPISTPNN 181

Qy        139 AFPQKTMDALSYGVF-------KGIYLLIKDAAQNK-KLYFTGGDGQFLANYFD-HAIYD 189
              : |  ::: |:|       :|:|  ||   : |  |  |||:   : :  :   |::
Db        182 LLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITGGNADLILSLIEIEFIFN 241

Qy        190 KLLIFRGMK 198
                |   |::
Db        242 IHLTVEGVR 250


RESULT 10
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

  Query Match             11.0%;  Score 120.5;  DB 9;  Length 248;
  Best Local Similarity   24.1%;  Pred. No. 0.0007;
  Matches   38;  Conservative   36;  Mismatches   45;  Indels   39;  Gaps    7;

Qy         68 YQGLGIDR----IAACYTIEDG-VVVDAGSAITIDII-SNSIHLGGFILPGIANYKKIYS 121
              || ||:||    :|| :  :   :|:| |:|:| |:: :: :||||:| ||:
Db         95 YQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMT------- 147

Qy        122 HISPRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK---- 167
                   ::|:  |        ||  :: : :|  |      | :|  |:::   : :
Db        148 ----LMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMA 203

Qy        168 --------KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                     ::: |||| : : :    |     |:| |:
Db        204 CELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVGL 241


RESULT 11
US-09-813-453A-13
; Sequence 13, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Synechocystis sp.
US-09-813-453A-13

  Query Match             10.7%;  Score 116.5;  DB 9;  Length 257;
  Best Local Similarity   30.4%;  Pred. No. 0.0018;
  Matches   52;  Conservative   26;  Mismatches   66;  Indels   27;  Gaps   11;

Qy         49 LKNQKNFINLEPYFLFDTIYQGLGIDRIAA--CYTIEDGVVVDAGSAITIDIISNSIHL- 105
                |||    :|| | |  |    |||         |      :||| |:|:||        |
Db         87 LKNLP-LVNLYPSFGIDRALAGLGTGLTYGFPC------LVVDGGTALTITGFDQDKKLV 139
```

```
Qy        106 GGFILPGIA-NYKKIYSHIS--PRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKD 162
              || ||||:      :   ::  |:|: :  |::  | :   |  |:  ||  |:   ::
Db        140 GGAILPGLGLQLATLGDRLAALPKLEMDQLTELP-DRWALDTPSAIFSGVVYGVLGALQS 198

Qy        163 AAQN-KKLY------FTGGDGQFLANYF-DHA-----IYDKLLIFRGMKKI 200
                |: :||:       |||||: |  :  :|:      :|  ||| ||  |
Db        199 YLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNLIFLGMAAI 249


RESULT 12
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
   CURRENT APPLICATION NUMBER: US/09/813,453A
   CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             10.5%;  Score 114.5;  DB 9;  Length 249;
  Best Local Similarity   26.6%;  Pred. No. 0.0027;
  Matches   42;  Conservative   28;  Mismatches   49;  Indels   39;  Gaps    8;

Qy         68 YQGLGIDR-IAA--CYTIEDG--VVVDAGSAITIDIIS-NSIHLGGFILPGIANYKKIYS 121
              || ||:|| :||   : :  |  :|:| |:|   | :| :  ||||:| ||:
Db         96 YQRLGMDRWLAALGAFHLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGM-------- 147

Qy        122 HISPRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK---- 167
                | ::|:  |        ||  :: : :|| |      | :|  |:::  |  :
Db        148 ---PLMRSQLRTHTRRIRYDDASAERALSSLSPGRSTVEAVERGCVLMLQGFAYTQLEQA 204

Qy        168 --------KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                     :: ||||   :      |     |:| |:
Db        205 RVLWGEEFTVFLTGGDAPLVRAALPQARVVPDLVFVGL 242


RESULT 13
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             10.4%;  Score 113.5;  DB 9;  Length 249;
  Best Local Similarity   26.5%;  Pred. No. 0.0033;
  Matches   41;  Conservative   26;  Mismatches   49;  Indels   39;  Gaps    7;

Qy         71 LGIDR---IAACYTIEDG--VVVDAGSAITIDII-SNSIHLGGFILPGIANYKKIYSHIS 124
              ||:||       | :  |  :|:| |:||| | : ::  ||||||:||:
Db         99 LGLDRWLAFVGAYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGM---------- 147

Qy        125 PRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK------- 167
              | ::::  |        ||  :| :  |  |      | :|  |:::  |  :
Db        148 PLMRNQLRTHTRRIRYDDAEAEKALVRLVPGRATAEAVERGCSLMLRGFAMTQIEIAREY 207

Qy        168 -----KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                  :: ||||   :|:     |     |:| |:
Db        208 WGDDFAIFVTGGDAVLVADVLPGARIVPDLVFVGL 242


RESULT 14
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
US-09-813-453A-45

  Query Match             10.3%;  Score 113;  DB 9;  Length 262;
  Best Local Similarity   21.8%;  Pred. No. 0.004;
  Matches   57;  Conservative   50;  Mismatches   93;  Indels   62;  Gaps   13;

Qy          1 MLLCDIGNSNANF-------------LDDNKYFT-----LNIDQFLE-----FKNEQKIF 37
              : : |:||:||               :: ::: |     : : | ||     |:: : |
Db          2 IFVLDVGNTNAVLGVFEEGELRQHWRMETDRHKTEDEYGMLVKQLLEHEGLSFEDVKGII 61

Qy         38 YINVNEHLKEHLKNQ-KNFINLEPYFLFDTIYQGL----------GIDRIAACYTIEDG- 85
                :|  :  |:  : :  ::|  :  |  ||           | |||       |
Db         62 VSSVVPPIMFALERMCEKYFKIKPLVVGPGIKTGLNIKYENPREVGADRIVNAVA---GI 118

Qy         86 -------VVVDAGSAITIDIISNSIH-LGGFILPGI-ANYKKIYSHISPRLKSEFNTQVS 136
                    ::|| |:| |   |:   | :|| | |||  : : :||  :   : |     |
Db        119 HLYGSPLIIVDFGTATTYCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRIEITKPSS 178

Qy        137 LDAFPQKTMDALSYGVFKG--------IYLLIKDAAQNKKLYFTGGDGQFL---ANYFDH 185
              :    : |: |:  |:  |         :  : ::| |  |:  |||  : :   :|  |
Db        179 V--VGKNTVSAMQSGILYGYVGQVEGIVKRMKEEAKQEPKVIATGGLAKLISEESNVID- 235

Qy        186 AIYDKLLIFRGMKKIIKENPNL 207
               : |  |  :|:  : : | ||
Db        236 -VVDPFLTLKGLYMLYERNANL 256


RESULT 15
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
```

```
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-813-453A-51

  Query Match           10.2%;  Score 111;  DB 9;  Length 260;
  Best Local Similarity  26.8%;  Pred. No. 0.0061;
  Matches   52;  Conservative   31;  Mismatches   64;  Indels   47;  Gaps   11;

Qy         48 HLKN-QKNFINLEPYFLFDTIYQGL----------GIDRI-----AACYTIEDGVVVDAG 91
              :|:|  : : |:||  : :   |:          | ||:     ||       ||:|:|
Db         72 NLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSG 131

Qy         92 SAITIDII-SNSIHLGGFILPGIANYKKIYSHIS----PRL---KSEFNTQVSLDAFPQK 143
              :| | ||: ::    || | ||| |      | :    ||:   :   |  |  |
Db        132 TATTFDIVAADGAFEGGIIAPGI-NLSMQALHEAAAKLPRIAIQRPAGNRIVGTD----- 185

Qy        144 TMDALSYGVFKGIYLLIK--------DAAQNKKLYFTGGDGQFLANYFDHAI-----YDK 190
              |: |:  ||| |   ||:        :  :   :  |||    :|: |: |      :|
Db        186 TVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGG----VASLFEGATDSIDHFDS 241

Qy        191 LLIFRGMKKIIKEN 204
               |  ||: :| : |
Db        242 DLTIRGLLEIYRRN 255


Search completed: June 24, 2003, 22:24:05
Job time : 10.5289 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 7.6231 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-21
Perfect score: 1093
Sequence: 1 MLLCDIGNSNANFLDDNKYF.....KLLIFRGMKKIIKENPNLLY 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1093	100.0	209	2	H81382	hypothetical prote
2	391.5	35.8	223	2	F64627	hypothetical prote
3	384.5	35.2	223	2	G71887	hypothetical prote
4	177.5	16.2	229	2	E70465	hypothetical prote
5	139.5	12.8	246	2	D72320	conserved hypothet
6	132	12.1	276	2	AI2292	hypothetical prote
7	126.5	11.6	224	2	A99571	conserved hypothet
8	121.5	11.1	262	2	F70165	conserved hypothet
9	120.5	11.0	248	2	H83111	hypothetical prote
10	116.5	10.7	257	2	S75559	hypothetical prote
11	111	10.2	261	2	B87489	transcription acti
12	110	10.1	254	2	F83660	hypothetical prote
13	105	9.6	3973	2	B71612	hypothetical prote
14	98.5	9.0	3119	2	T18414	protein g377 - mal
15	98	9.0	259	2	AF1102	conserved hypothet
16	97	8.9	967	2	A64710	type III restricti
17	96	8.8	356	2	B70424	lipid A disacchari
18	95	8.7	969	2	E71810	type III restricti
19	94.5	8.6	274	2	H86937	conserved hypothet
20	94.5	8.6	481	2	A84130	two-component sens
21	94	8.6	465	2	A81419	hypothetical prote
22	93.5	8.6	381	2	T28378	ORF MSV217 SCG gen
23	93	8.5	259	2	AF1464	conserved hypothet
24	92.5	8.5	2652	1	VFIHB2	genome polyprotein
25	91.5	8.4	272	2	A70955	hypothetical prote
26	90	8.2	1272	2	S60999	ubiquitin-specific
27	89.5	8.2	242	2	A82637	conserved hypothet
28	89.5	8.2	256	2	A95100	ABC transporter, A
29	89.5	8.2	263	2	C97968	hypothetical prote
30	89	8.1	622	2	E97126	DNA mismatch repai
31	89	8.1	1024	2	C64208	hypothetical prote
32	88.5	8.1	431	2	D86433	hypothetical prote
33	88.5	8.1	592	2	B81009	BirA protein/Bvg a
34	88.5	8.1	901	2	D70116	transcription elon
35	88.5	8.1	1473	2	A35186	salivary agglutini
36	87.5	8.0	1161	2	T18400	glutamate-cysteine
37	87.5	8.0	1307	2	T25563	hypothetical prote
38	87	8.0	439	2	T28196	hypothetical prote
39	87	8.0	713	2	A56268	Fe-regulated prote
40	86.5	7.9	243	2	B86740	lipopolysaccharide
41	86.5	7.9	448	2	G81382	probable oxidoredu
42	86	7.9	344	2	G70458	hypothetical prote
43	85.5	7.8	276	2	S70995	cell surface adhes
44	85.5	7.8	279	2	G97246	araC-type DNA-bind
45	85.5	7.8	356	2	D97732	hypothetical prote

ALIGNMENTS

RESULT 1
H81382
hypothetical protein Cj0394c [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: H81382
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Ch
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: H81382
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74230.1; PID:g69
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0394c
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 100.0%; Score 1093; DB 2; Length 209;
Best Local Similarity 100.0%; Pred. No. 5.3e-83;
Matches 209; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60

Qy         61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120

Qy        121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180

Qy        181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
              |||||||||||||||||||||||||||||
Db        181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
```

RESULT 2
F64627
hypothetical protein HP0862 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
C;Accession: F64627
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann,
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; Mc
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey
Nature 388, 539-547, 1997

```
A;Authors: Wallin, H.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64627
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <TOM>
A;Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07916.1; PID:g231400
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

  Query Match             35.8%;  Score 391.5;  DB 2;  Length 223;
  Best Local Similarity   39.2%;  Pred. No. 4.2e-25;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 3
G71887
hypothetical protein jhp0796 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Aug-2000
C;Accession: G71887
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: G71887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <ARN>
A;Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g415536
A;Experimental source: strain J99
C;Genetics:
Gene: jhp0796
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

  Query Match             35.2%;  Score 384.5;  DB 2;  Length 223;
  Best Local Similarity   38.2%;  Pred. No. 1.6e-24;
  Matches   78;  Conservative   46;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LILCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :||:|||||||||::    |||| ||||:| |
Db         73 GFFHLETDYIGLGIDRQMACLAVVNGVIVDAGSAITIDLVKEGKHLGGCILPGLAQYVHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |:  |    ||:  |: | ||::||:   |   |:  |:::|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCIQHLAKDQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 4
E70465
hypothetical protein aq_1924 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: E70465
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-229 <AQF>
A;Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AEC
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1924

  Query Match             16.2%;  Score 177.5;  DB 2;  Length 229;
  Best Local Similarity   25.8%;  Pred. No. 1.9e-07;
  Matches   62;  Conservative   44;  Mismatches   83;  Indels   51;  Gaps   10;

Qy          2 LLCDIGNSNANF-LDDNKYFTLNIDQFLEFKNEQ---------KIFYINVNEHLKEHLK- 50
              |  |:|||: :  | : |     :  ||:  :|:          |   |:| :   | ::
Db          4 LTVDVGNSSVDIALWEGK----KVKDFLKLSHEEFLKEEFPKLKALGISVKQSFSEKVRG 59

Qy         51 --------NQKNFINLEPYFLFDTIYQGLGIDRIAACYTI-----EDGVVVDAGSAITID 97
                     ::||     |   :|    || ||:|  |:       :: ||: ||:|: ||
Db         60 KIPKIKFLKKENFPIQVDYKTPET----LGTDRVALAYSAKKFYGKNVVVISAGTALVID 115

Qy         98 IISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIY 157
              ::      ||||  |:    || | ::  :   |  :| :    : | : :  | ::
Db        116 LVLEGKFKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEI-FLGRSTRECVLGGAYREST 174

Qy        158 LLIKDA--------AQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
                ||            :  |:  |||:|:: : :    ||| ||: |||:       ||||
Db        175 EFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGMR-------NLLY 224


RESULT 5
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hi
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g49
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             12.8%;  Score 139.5;  DB 2;  Length 246;
  Best Local Similarity   22.7%;  Pred. No. 0.00029;
  Matches   58;  Conservative   38;  Mismatches   82;  Indels   77;  Gaps   10;

Qy          2 LLCDIGNSNANF--LDDNKYF---------------------TLNIDQFLEFKN------ 32
              || |:||::: |   :| | |                       |  |   | |
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         33 ----------EQKIFYIN------VNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDR 75
```

```
                    || |:|:       |  :| ::||                :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNVKNPSE---------------VGADR 106

Qy         76 IAACYTI-----EDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSE 130
              :|          ::|:::| |:| |:|:: |  : || ||||   :  ::|       |
Db        107 VANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGF--FMMVHSLFRGTAKLP 164

Qy        131 FNTQVSLDAFPQK-TMDALSYGVFKG-IYLL------IKDAAQNKKLYFTGGDGQFLANY 182
                  |    | | : :  ||  | :| |      ||:   :  :  |||  : : :
Db        165 LVEVKPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDM 224

Qy        183 FDHAIYDKLLIFRGM 197
                | |:|: |  :|:
Db        225 IKHEIFDEDLTIKGV 239


RESULT 6
AI2292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
;Species: Nostoc sp.
;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

  Query Match             12.1%;  Score 132;  DB 2;  Length 276;
  Best Local Similarity   25.8%;  Pred. No. 0.0014;
  Matches   69;  Conservative   38;  Mismatches   84;  Indels   76;  Gaps   14;

Qy          2 LLCDIGNSN---ANFLDDNKYFTLNIDQF-------------LEFKNEQK-IFY----- 38
              |  :|||.    | |: ::  || : :                ||  :|:| ||:
Db         15 LALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEVGSEEKEIFFTFFPL 74

Qy         39 --------INVNEHLKEHLKNQKNFINLEPYFL----FDTIYQGLGIDRIAACYTIEDG- 85
                    : :  : :     :|::|:    |     : ||  ||||| | :
Db         75 PPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPTLGIDRALALWGAGMSW 134

Qy         86 ----VVVDAGSAITIDIISNSIHL-GGFILPGIANYKKIYSHISPRLKSEFNTQVSLDA- 139
                 :|:|||:|:|        :| || ||||:          : :|      || ::|
Db        135 GFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQL-----PQVEMEAI 189

Qy        140 ------FPQKTMDALSYGVFKGIYLLI---KDAAQ-------NKKLYFTGGDGQFLANYF 183
                   | | :|:  ||   || ||   :|  :        : |:   |||   | ||
Db        190 KSLPPRFALNTTEAIQSGV---IYTLIAGMRDFTEEWLSLFPDGKVAIKGGDRILLLNYL 246

Qy        184 DHAIYDKL---------LIFRGMKKII 201
                |:|  |         ||| ||: |:
Db        247 -QALYPDLAARLIVEPNLIFWGMQTIV 272


RESULT 7
A99571
conserved hypothetical protein. MYPU_4730 [imported] - Mycoplasma pulmonis (strain UAB C
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: A99571
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: A99571
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-224 <KUR>
A;Cross-references: GB:AL445566; PID:g14089887; PIDN:CAC13646.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_4730
A;Genetic code: SGC3

  Query Match             11.6%;  Score 126.5;  DB 2;  Length 224;
  Best Local Similarity   21.8%;  Pred. No. 0.003;
  Matches   41;  Conservative   40;  Mismatches   66;  Indels   41;  Gaps    5;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLE------FKNEQKIFYINVNEHLKEHLKNQKN 54
              :|:| : |:  |             | ||      |||  :  |:|: :: | :  : :
Db         56 ILICSVRNAKEN------------QILEFKLKSIFKNAKIDFFIHKSQSLVKFCQKE-- 100

Qy         55 FINLEPYFLFDTIYQGLGIDRIAACYTI----EDGVVVDAGSAITIDIISNSIHLGGFIL 110
                        :   :|:| :|  | :    :: : :  |:|| |  | ||   |  |
Db        101 -----------LTSEIGLDIVANAYYVLHKSKNAIFISLGTAIVITQIKNSSIEGISIY 148

Qy        111 PGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIK----DAAQN 166
              |||    | : ::  :::| ||       : |::::|  : :|   |:|    :  :
Db        149 PGIYQSFKNFFNVVAKIESNFNIIKIPPILGKTTLESISSSLVRGSVFLLKGYIDEIDKT 208

Qy        167 KKLYFTGG 174
                :: |||
Db        209 SDIFITGG 216


RESULT 8
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: F70165
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.;
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch,
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: F70165
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-262 <KLE>
A;Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g2
A;Experimental source: strain B31

  Query Match             11.1%;  Score 121.5;  DB 2;  Length 262;
  Best Local Similarity   24.9%;  Pred. No. 0.0095;
  Matches   62;  Conservative   42;  Mismatches   86;  Indels   59;  Gaps   14;

Qy          2 LLCDIGNSNANF--LDDNKYFTLNIDQFLEFK--------------------NEQKIFYI 39
              |: ||||::  |    ||:     :: |:: |                    |  |:|
Db          9 LIIDIGNTSIAFALFKDNQ-----VNLFIKMKTNLMLRYDEVYSFFEENFDFNVNKVFIS 63

Qy         40 NVNEHLKEHLKNQ-KNFINLEPYFL-FDTIYQ-----------GLGID----RIAAC--Y 80
              :|    | |  ||   :|  ::| |: || |             |||  |     :||   |
Db         64 SVVPILNETFKNVIFSFFKIKPLFIGFDLNYDLTFNPYKSDKFLLGSDVFANLVAAIENY 123

Qy         81 TIEDGVVVDAGSAITIDIISNSIH-LGGFILPG-IANYKKIYSHISPRLKSEFNTQVSLD 138
              : |: :||| |:| ||  :|     ||| |  | : |:  :    :  |  :|      :
Db        124 SFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLD--NAYLIKKFPISTPNN 181

Qy        139 AFPQKTMDALSYGVF-------KGIYLLIKDAAQNK-KLYFTGGDGQFLANYFD-HAIYD 189
                 : |  ::: |:|       :|:|  ||   : |  |  |||:   : :  :   |::
Db        182 LLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITGGNADLILSLIEIEFIFN 241

Qy        190 KLLIFRGMK 198
                |   |::
Db        242 IHLTVEGVR 250
```

```
RESULT 9
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

  Query Match             11.0%;  Score 120.5;  DB 2;  Length 248;
  Best Local Similarity   24.1%;  Pred. No. 0.011;
  Matches   38;  Conservative   36;  Mismatches   45;  Indels   39;  Gaps    7;

Qy         68 YQGLGIDR----IAACYTIEDG-VVVDAGSAITIDII-SNSIHLGGFILPGIANYKKIYS 121
              || ||:||    :|| :  :   :|:| |:|:| |:: :: :||||:| ||:
Db         95 YQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMT------- 147

Qy        122 HISPRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK---- 167
                 ::|:  |         ||  :: : :|  |      | :|  |:::   : :
Db        148 ----LMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMA 203

Qy        168 --------KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                     ::: |||| : : :    |     |:| |:
Db        204 CELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVGL 241


RESULT 10
S75559
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75559
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
A Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75559
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-257 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18120.1; PID:d101885
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

  Query Match             10.7%;  Score 116.5;  DB 2;  Length 257;
  Best Local Similarity   30.4%;  Pred. No. 0.024;
  Matches   52;  Conservative   26;  Mismatches   66;  Indels   27;  Gaps   11;

Qy         49 LKNQKNFINLEPYFLFDTIYQGLGIDRIAA--CYTIEDGVVVDAGSAITIDIISNSIHL- 105
              |||    :|| | |  |    |||        |      :||| |:|:||        |
Db         87 LKNLP-LVNLYPSFGIDRALAGLGTGLTYGFPC------LVVDGGTALTITGFDQDKKLV 139

Qy        106 GGFILPGIA-NYKKIYSHIS--PRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKD 162
              || ||||:      :   ::  |:|: :  |::  | :   |  |:  ||  |:   ::
Db        140 GGAILPGLGLQLATLGDRLAALPKLEMDQLTELP-DRWALDTPSAIFSGVVYGVLGALQS 198

Qy        163 AAQN-KKLY------FTGGDGQFLANYF-DHA-----IYDKLLIFRGMKKI 200
                |: :||:       |||||: |  :  :|:      :|  ||| ||  |
Db        199 YLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNLIFLGMAAI 249


RESULT 11
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; K
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             10.2%;  Score 111;  DB 2;  Length 261;
  Best Local Similarity   26.8%;  Pred. No. 0.07;
  Matches   52;  Conservative   31;  Mismatches   64;  Indels   47;  Gaps   11;

Qy         48 HLKN-QKNFINLEPYFLFDTIYQGL----------GIDRI-----AACYTIEDGVVVDAG 91
              :|:|  : : |:||  : :    |:              | ||:     ||       ||:|:|
Db         73 NLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSG 132

Qy         92 SAITIDII-SNSIHLGGFILPGIANYKKIYSHIS----PRL---KSEFNTQVSLDAFPQK 143
              :| | ||: ::    || | ||| |      | :    ||:   :   |  |  |
Db        133 TATTFDIVAADGAFEGGIIAPGI-NLSMQALHEAAAKLPRIAIQRPAGNRIVGTD----- 186

Qy        144 TMDALSYGVFKGIYLLIK--------DAAQNKKLYFTGGDGQFLANYFDHAI-----YDK 190
              |: |:  ||| |   ||:        :  :   :  |||    :|: |: |      :|
Db        187 TVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGG----VASLFEGATDSIDHFDS 242

Qy        191 LLIFRGMKKIIKEN 204
               |  ||: :| : |
Db        243 DLTIRGLLEIYRRN 256


RESULT 12
F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83660
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.;
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03805.1; GSPDB:
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0086
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             10.1%;  Score 110;  DB 2;  Length 254;
  Best Local Similarity   23.4%;  Pred. No. 0.081;
  Matches   60;  Conservative   47;  Mismatches   93;  Indels   56;  Gaps   14;

Qy          1 MLLCDIGNSN---ANFLDDNKYFTLNI--------DQF------------LEFKNEQKIF 37
```

```
              :|: |:||:|     : |:       :        |::            |:|::   |
Db          2 ILVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGIV 61

Qy         38 YINVNEHLKEHLKNQ-KNFINLEPYFLFDTIYQGL----------GIDRIA-ACYTIE-- 83
               :|    :   |:   | : :: |  :   |  ||         | |||  |   ||
Db         62 ISSVVPPMMFSLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRIVNAVAAIELY 121

Qy         84 --DGVVVDAGSAITIDIIS-NSIHLGGFILPGI-ANYKKIYSHIS--PRLKSEFNTQVSL 137
                 :||| |:| |  :|:    : || | |||  : : :|   |  ||::     ||
Db        122 GYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQV-- 179

Qy        138 DAFPQKTMDALSYGVF-------KGIYLLIKDAAQNK-KLYFTGGDGQFLANYFDHA-IY 188
                  |:|::  |:|        |:   :|  |::: |:  |||  : :    :   :
Db        180 --VGTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVI 237

Qy        189 DKLLIFRGMKKIIKEN 204
              |  |  :|:: | |:|
Db        238 DSFLTLKGLQLIYKKN 253


RESULT 13
B71612
hypothetical protein PFB0555c - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C;Accession: B71612
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71612
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-3973 <GAR>
A;Cross-references: GB:AE001402; GB:AE001362; NID:g3845213; PIDN:AAC71900.1; PID:g384521
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0555c
C;Superfamily: Plasmodium falciparum hypothetical protein PFB0555c

  Query Match             9.6%;  Score 105;  DB 2;  Length 3973;
  Best Local Similarity   22.1%;  Pred. No. 6.4;
  Matches   55;  Conservative   42;  Mismatches   88;  Indels   64;  Gaps   11;

Qy          8 NSNANFLDDNKYFTLNIDQFLEFKNEQKI----------------------FYIN-VNE 43
              |:| :  ||: |   |:    : :|  ||                      |:| : |
Db       3066 NNNGDGYDDDGYGEKNVSGIYK-ENNNKINVKGDIYNIDNINVYPLNGKLVSIYLNTLKE 3124

Qy         44 HLKE----HLKNQKNFINLEPYFLFDTIYQGLGIDRIAACYTI--------------EDG 85
                |||    || : ||  |:|  |  : :      :||   | :                :
Db       3125 ILKECYECHLGHMKNNENMEKSFFIEHLLLYFLYNRINTIYELFYNFYFTYLRKKENNND 3184

Qy         86 VVVDAGSAITIDIISNSIH-----LGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAF 140
              :::|  :    ::| | |:      :  |    : : :  | :|:         | |: :
Db       3185 ILLDIVNEHIYNLIGNKIYDQINKINNF----LDDKQYYYFYINTLTFITLNKQICIYII 3240

Qy        141 PQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKI 200
               :| :: | |  |  ||  : |  :|    :   : |:: |  || |:  |::|     |
Db       3241 KKKILNKLIYIPF--IYHSLFDKNKNFTSIYHINNNQYIRNK-DHIIFCSLIVF-----I 3292

Qy        201 IKENPNLLY 209
              ||    |||
Db       3293 IK----LLY 3297


RESULT 14
T18414
protein g377 - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C;Accession: T18414
R;Handman, E.; Osborn, A.H.; Symons, F.; Van Driel, R.; Cappai, R.
Mol. Biochem. Parasitol. 74, 143-156, 1995
A;Title: The Leishmania promastigote surface antigen 2 complex is differentially ex
A;Reference number: Z18933; MUID:96360472; PMID:8719156
A;Accession: T18414
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3119 <HAN>
A;Cross-references: EMBL:L04161; NID:g309687; PID:g309688; PIDN:AAC37257.1

  Query Match             9.0%;  Score 98.5;  DB 2;  Length 3119;
  Best Local Similarity   20.2%;  Pred. No. 16;
  Matches   51;  Conservative   41;  Mismatches   64;  Indels   97;  Gaps   11;

Qy         18 KYFTLNIDQFLEFKNEQKIFYINVNEHL-KEHLKNQKN----------FINLEPYFLFD 65
              ::| :  || :   ::  ||  |: | : ||:| ::|           :::|| :|
Db       1662 RHFVMTFDQKIYDYDDHLIFVYNIKELIYKENLADEKYNANYYYEMNRFYLHLEEFFYIL 1721

Qy         66 TIY------QGLGIDRIAACYTIEDGVVVDAGSAITI---------------------- 96
                    | ||     |  : |: :    || | :
Db       1722 KHYVELRKIQDLG---EVAIPSFEENIKT-KGSLIDVHQKDKDGTQIIEYLERNVLHIYS 1777

Qy         97 ---DIISNSIHLGGFILPGIAN-------YKKIYSHISPRLKSEFNTQVSLDAFPQKTMD 146
                 :| :|:|||  :::| ::|       :||: :   | |
Db       1778 ILRNIKNNNIHLYHYLVPDLSNILIDPEEFKKMSNGEQPELNK----------------- 1820

Qy        147 ALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL----------ANYFDHAIYDKLLIFRG 196
                       ::  : ||  | |      :|| :           |:|:|  ||||| ::
Db       1821 ---------LHKYLLDAYTNLK-----KEGQVVDQIYVEYINFHNHFEHT-YDKLLWYKT 1865

Qy        197 MKKIIKENPNLLY 209
              : |   |    :|
Db       1866 ILKRYIERSKKIY 1878


RESULT 15
AF1102
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1102
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blo
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsih
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Weh
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1102
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0221
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             9.0%;  Score 98;  DB 2;  Length 259;
  Best Local Similarity   21.5%;  Pred. No. 0.82;
  Matches   55;  Conservative   48;  Mismatches   97;  Indels   56;  Gaps   13;

Qy          1 MLLCDIGNSNANF-------------LDDNKYFT-----LNIDQFLEFKN-----EQKIF 37
              :|: |:||:|                :  ::: |     : :  |  : |     | |
Db          2 ILVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGII 61

Qy         38 YINVNEHLKEHLKNQ-KNFINLEPYFLFDTIYQGL----------GIDRI--AACYTIED 84
               :|    :   ::     : |: |  :   |  ||          | |||  |   : |
Db         62 ISSVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEEY 121

Qy         85 G---VVVDAGSAITIDIISNS-IHLGGFILPGI-ANYKKIYSHIS--PRLKSEFNTQVSL 137
              |   :||| |:| |    |  | :: || | |||  : : :|:  :  ||:   ::|:
Db        122 GTPVIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQI-- 179
```

```
Qy        138 DAFPQKTMDALSYGVFKG--------IYLLIKDAAQNKKLYFTGGDGQFLANYFDHA-IY 188
                  : |: ::  |:| |        |  : | :  :  :  |||  : :        |
Db        180 --IGKSTVSSMQAGIFYGFVGQCEGIIAEMKKQSNASPVVVATGGLARMITEKSSAVDIL 237

Qy        189 DKLLIFRGMKKIIKEN 204
              |  |  :|:: : : |
Db        238 DPFLTLKGLELLYRRN 253
```

Search completed: June 24, 2003, 22:03:19
Job time : 9.6231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 21.4576 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-21
Perfect score: 1093
Sequence: 1 MLLCDIGNSNANFLDDNKYF.....KLLIFRGMKKIIKENPNLLY 209

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1093	100.0	209	23	AAU91165	Pantothenate kinas
2	391.5	35.8	223	19	AAW98422	H. pylori GHPO 344
3	391.5	35.8	223	23	AAU91161	Helicobacter pylor
4	391.5	35.8	223	23	AAU91181	Pantothenate kinas
5	177.5	16.2	229	23	AAU91159	Aquifex aeolicus p
6	139.5	12.8	246	23	AAU91156	Thermotoga maritim
7	126.5	11.6	244	23	AAU91168	Pantothenate kinas
8	124	11.3	241	23	AAU91179	Pantothenate kinas
9	121.5	11.1	262	23	AAU91158	Borrelia burgdorfe
10	120.5	11.0	248	23	AAU91164	Pantothenate kinas
11	116.5	10.7	257	23	AAU91160	Synechocystis pant
12	114.5	10.5	249	23	AAU91178	Pantothenate kinas
13	113.5	10.4	249	23	AAU91182	Pantothenate kinas
14	113	10.3	262	23	AAU91170	Pantothenate kinas
15	111	10.2	260	23	AAU91173	Pantothenate kinas
16	110	10.1	254	23	AAU91171	Pantothenate kinas
17	109	10.0	258	22	AAU01243	B. subtilis novel
18	109	10.0	258	23	AAU91149	Bacillus subtilis
19	108	9.9	257	23	AAU91174	Pantothenate kinas
20	105	9.6	3973	21	AAB18253	Plasmodium falcipa
21	98	9.0	259	23	ABB47661	Listeria monocytog
22	96	8.8	258	23	AAU91153	Rhodobacter capsul
23	91.5	8.4	272	22	AAG81225	Mycobacterium tube
24	91.5	8.4	272	23	AAU91152	Mycobacterium tube
25	89.5	8.2	242	23	AAU91180	Pantothenate kinas
26	89.5	8.2	256	22	AAU37733	Streptococcus pneu
27	89.5	8.2	257	21	AAY81593	Streptococcus pneu
28	89	8.1	1024	18	AAW19604	Mycoplasma genital
29	88.5	8.1	431	23	ABB91199	Herbicidally activ
30	88.5	8.1	592	20	AAY38615	Neisseria meningit
31	88.5	8.1	592	21	AAY74912	Neisseria meningit
32	88.5	8.1	592	23	AAU91169	Pantothenate kinas
33	88.5	8.1	1264	14	AAR36730	Ubiquitin-specific
34	87.5	8.0	389	21	AAY74909	Neisseria meningit
35	87	8.0	713	17	AAW05180	Neisseria gonorrho
36	87	8.0	713	22	AAE05407	N. gonorrheae or N
37	86.5	7.9	219	23	AAU91176	Pantothenate kinas
38	86.5	7.9	243	23	ABB54238	Lactococcus lactis
39	86.5	7.9	754	23	ABP28632	Streptococcus poly
40	85.5	7.8	256	23	ABP28305	Streptococcus poly
41	85	7.8	212	23	AAU91177	Pantothenate kinas
42	85	7.8	318	23	ABP26545	Streptococcus poly
43	85	7.8	455	21	AAY74910	Neisseria meningit
44	85	7.8	592	20	AAY38616	Neisseria meningit
45	85	7.8	592	21	AAY74913	Neisseria meningit

ALIGNMENTS

RESULT 1
AAU91165
ID AAU91165 standard; Protein; 209 AA.
XX
AC AAU91165;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #3.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Campylobacter jejuni.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -

```
XX
PS   Claim 6; Page 84-85; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   209 AA;

  Query Match            100.0%;  Score 1093;  DB 23;  Length 209;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-112;
  Matches  209;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKNQKNFINLEP 60

Qy         61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIY 120

Qy        121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFLA 180

Qy        181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
              |||||||||||||||||||||||||||||
Db        181 NYFDHAIYDKLLIFRGMKKIIKENPNLLY 209


RESULT 2
AAW98422
ID   AAW98422 standard; Protein; 223 AA.
XX
AC   AAW98422;
XX
DT   31-MAR-1999  (first entry)
XX
DE   H. pylori GHPO 344 protein.
XX
KW   GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW   peptic ulcer disease.
XX
OS   Helicobacter pylori.
XX
PN   WO9843478-A1.
XX
PD   08-OCT-1998.
XX
PF   01-APR-1998;   98WO-US06371.
XX
PR   29-JUL-1997;   97US-0902615.
PR   01-APR-1997;   97US-0833457.
PR   24-JUN-1997;   97US-0881227.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI   Al-Garawi A,  Kleanthous H,  Miller C,  Oomen RP,  Tomb J;
XX
DR   WPI; 1998-542293/46.
DR   N-PSDB; AAX14141.
XX
PT   New isolated Helicobacter polynucleotides - used to develop products
PT   for the diagnosis, prevention and treatment of Helicobacter
PT   infections and gastrointestinal diseases
XX
PS   Claim 8; Page 709-710; 2054pp; English.
XX
CC   This sequence represents a Helicobacter pylori GHPO protein of the
CC   invention. The polypeptides can be used for preventing or treating
CC   Helicobacter infections, and gastroduodenal diseases associated with
CC   these infections, including acute, chronic, and atrophic gastritis, and
CC   peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC   used for the production of antibodies. The products can also be used for
CC   detection and diagnosis.
XX
SQ   Sequence   223 AA;

  Query Match            35.8%;  Score 391.5;  DB 19;  Length 223;
  Best Local Similarity  39.2%;  Pred. No. 7.6e-35;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::|||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
                :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 3
AAU91161
ID   AAU91161 standard; Protein; 223 AA.
XX
AC   AAU91161;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Helicobacter pylori pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Helicobacter pylori.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 78-79; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
```

```
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   223 AA;

  Query Match             35.8%;  Score 391.5;  DB 23;  Length 223;
  Best Local Similarity   39.2%;  Pred. No. 7.6e-35;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |: |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 4
AAU91181
ID   AAU91181 standard; Protein; 223 AA.
XX
AC   AAU91181;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #19.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Helicobacter pylori.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54202.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 112-113; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   223 AA;

  Query Match             35.8%;  Score 391.5;  DB 23;  Length 223;
  Best Local Similarity   39.2%;  Pred. No. 7.6e-35;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59
              ::||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72

Qy         60 PYFLFDTIYQQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132

Qy        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179
              |   :  |: |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192

Qy        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
              : :  |::  : |:| ||:  :|:
Db        193 SAFLPHSVCKERLVFDGMEIALKK 216


RESULT 5
AAU91159
ID   AAU91159 standard; Protein; 229 AA.
XX
AC   AAU91159;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Aquifex aeolicus pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Aquifex aeolicus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 77; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
```

```
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   229 AA;

  Query Match             16.2%;  Score 177.5;  DB 23;  Length 229;
  Best Local Similarity   25.8%;  Pred. No. 3.7e-11;
  Matches   62;  Conservative   44;  Mismatches   83;  Indels   51;  Gaps   10;

Qy          2 LLCDIGNSNANF-LDDNKYFTLNIDQFLEFKNEQ---------KIFYINVNEHLKEHLK- 50
              |  |:|||: :  | : |     :  ||:  :|:         |   |:| :   | ::
Db          4 LTVDVGNSSVDIALWEGK----KVKDFLKLSHEEFLKEEFPKLKALGISVKQSFSEKVRG 59

Qy         51 --------NQKNFINLEPYFLFDTIYQGLGIDRIAACYTI-----EDGVVVDAGSAITID 97
                    ::||     |   :|    || ||:|  |:       :: ||: ||:|: ||
Db         60 KIPKIKFLKKENFPIQVDYKTPET----LGTDRVALAYSAKKFYGKNVVVISAGTALVID 115

Qy         98 IISNSIHLGGFILPGIANYKKIYSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIY 157
              ::      ||||  |:    || | ::  :   |  :| :    : | : :  | ::
Db        116 LVLEGKFKGGFITLGLGKKLKILSDLAEGIPEFFPEEVEI-FLGRSTRECVLGGAYREST 174

Qy        158 LLIKDA--------AQNKKLYFTGGDGQFLANYFDHAIYDKLLIFRGMKKIIKENPNLLY 209
               ||           :  |:  |||:|:: : :    ||| ||: |||:       ||||
Db        175 EFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGMR-------NLLY 224


RESULT 6
AAU91156
ID   AAU91156 standard; Protein; 246 AA.
XX
AC   AAU91156;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Thermotoga maritima pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Thermotoga maritima.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 74-75; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   246 AA;

  Query Match             12.8%;  Score 139.5;  DB 23;  Length 246;
  Best Local Similarity   22.7%;  Pred. No. 6.6e-07;
  Matches   58;  Conservative   38;  Mismatches   82;  Indels   77;  Gaps   10;

Qy          2 LLCDIGNSNANF--LDDNKYF--------------------TLNIDQFLEFKN------ 32
              || |:||::: |   :| | |                     |  |   | |
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         33 ----------EQKIFYIN------VNEHLKEHLKNQKNFINLEPYFLFDTIYQGLGIDR 75
                       || |:|:       |  :| ::||                    :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKNGCVKWNVKNPSE---------------VGADR 106

Qy         76 IAACYTI-----EDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSHISPRLKSE 130
              :|          ::|:::| |:| |:|:: |  : || ||||   :  ::|      |
Db        107 VANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGF--FMMVHSLFRGTAKLP 164

Qy        131 FNTQVSLDAFPQK-TMDALSYGVFKG-IYLL------IKDAAQNKKLYFTGGDGQFLANY 182
                    |   | | : :  ||  | :| |      ||:   :  :  |||  .: : :
Db        165 LVEVKPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDM 224

Qy        183 FDHAIYDKLLIFRGM 197
               | |:|: |  :|:
Db        225 IKHEIFDEDLTIKGV 239


RESULT 7
AAU91168
ID   AAU91168 standard; Protein; 244 AA.
XX
AC   AAU91168;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #6.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54189.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 8; Page 94-95; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
```

```
CC   described in the invention.
XX
SQ   Sequence   244 AA;

  Query Match             11.6%;  Score 126.5;  DB 23;  Length 244;
  Best Local Similarity   28.0%;  Pred. No. 1.8e-05;
  Matches   45;  Conservative   29;  Mismatches   50;  Indels   37;  Gaps    8;

Qy         71 LGIDRIAA------CYTIEDGVVVDAGSAITIDIIS-NSIHLGGFILPGI-ANYKKIY-- 120
              || ||:||       |   : :|:|||:||| : :|   |:||| | ||:   :| ::
Db         93 LGADRLAAVVGAHSLYPNTELLVIDAGTAITYERVSAEGIYLGGNISPGLHLRFKALHLF 152

Qy        121 ---------SHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGI------YLLIKDAAQ 165
                      | |||::           :   | :|:: ||  |:      |:    | :
Db        153 TGRLPLIDPSGISPKIAE----------YGSSTEEAITAGVIHGLAGEIDRYIDDLHAKE 202

Qy        166 NKK-LYFTGGDGQFLANYFDHAI-YDKLLIFRGMKKIIKEN 204
              :  :  ||||  :||      |     |:  |: :|:: |
Db        203 GRSAVILTGGDANYLARIIRSGILIHPDLVLLGLNRILEYN 243


RESULT 8
AAU91179
ID   AAU91179 standard; Protein; 241 AA.
XX
AC   AAU91179;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #17.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Thiobacillus ferrooxidans.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54200.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 109-110; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   241 AA;

  Query Match             11.3%;  Score 124;  DB 23;  Length 241;
  Best Local Similarity   28.4%;  Pred. No. 3.3e-05;
  Matches   40;  Conservative   22;  Mismatches   65;  Indels   14;  Gaps    4;

Qy         69 QGLGIDR----IAAC--YTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKIYSH 122
              : || ||    :||   |  :| :|:| |:|||||:::     || ||||||   :   |
Db         95 ESLGFDRRCCLLAAAMDYPGQDSIVIDMGTAITIDLLAGGHFRGGRILPGIAMSLR-GLH 153

Qy        123 ISPRLKSEFNTQVSLDAFPQKTMDALSYGV-------FKGIYLLIKDAAQNKKLYFTGGD 175
                | |       :     | :|:  ||        :|     :  :   ::  ||||
Db        154 EGTALLPEVVLNAPAEMLGNDTSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGD 213

Qy        176 GQFLANYFDHAIYDKLLIFRG 196
              :        ::|   |: ||
Db        214 AERWQPGIAGSLYQPHLLLRG 234


RESULT 9
AAU91158
ID   AAU91158 standard; Protein; 262 AA.
XX
AC   AAU91158;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Borrelia burgdorferi pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Borrelia burgdorferi.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 76; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   262 AA;

  Query Match             11.1%;  Score 121.5;  DB 23;  Length 262;
  Best Local Similarity   24.9%;  Pred. No. 7.1e-05;
  Matches   62;  Conservative   42;  Mismatches   86;  Indels   59;  Gaps   14;

Qy          2 LLCDIGNSNANF--LDDNKYFTLNIDQFLEFK-------------------NEQKIFYI 39
              |: ||||::  |    ||:     :: |:: |                   |  |:|
Db          9 LIIDIGNTSIAFALFKDNQ-----VNLFIKMKTNLMLRYDEVYSFFEENFDFNVNKVFIS 63

Qy         40 NVNEHLKEHLKNQ-KNFINLEPYFL-FDTIYQ-----------GLGID----RIAAC--Y 80
```

```
           :|   | |  ||   :|  ::| |: ||  |            || |     :||   |
Db          64 SVVPILNETFKNVIFSFFKIKPLFIGFDLNYDLTFNPYKSDKFLLGSDVFANLVAAIENY 123

Qy          81 TIEDGVVVDAGSAITIDIISNSIH-LGGFILPG-IANYKKIYSHISPRLKSEFNTQVSLD 138
           : |: :||| |:| ||  :|     ||| |  | : |:  :    :  |  :|      :
Db         124 SFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLD--NAYLIKKFPISTPNN 181

Qy         139 AFPQKTMDALSYGVF-------KGIYLLIKDAAQNK-KLYFTGGDGQFLANYFD-HAIYD 189
              : |  ::: |:|       :|:|  ||   : |  | |||:   : :  :   |::
Db         182 LLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITGGNADLILSLIEIEFIFN 241

Qy         190 KLLIFRGMK 198
                |   |::
Db         242 IHLTVEGVR 250


RESULT 10
AAU91164
ID   AAU91164 standard; Protein; 248 AA.
XX
AC   AAU91164;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #2.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 83-84; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   248 AA;

  Query Match             11.0%;  Score 120.5;  DB 23;  Length 248;
  Best Local Similarity   24.1%;  Pred. No. 8.5e-05;
  Matches   38;  Conservative   36;  Mismatches   45;  Indels   39;  Gaps    7;

Qy          68 YQGLGIDR----IAACYTIEDG-VVVDAGSAITIDII-SNSIHLGGFILPGIANYKKIYS 121
               || ||:||    :|| :  :   :|:| |:|:| |:: :: :||||:| ||:
Db          95 YQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMT------- 147

Qy         122 HISPRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK---- 167
                  ::|:  |       ||  :: : :|  |      | :|  |:::   : :
Db         148 ----LMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMA 203

Qy         168 --------KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                     ::: |||| : : :    |     |:| |:
Db         204 CELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVGL 241


RESULT 11
AAU91160
ID   AAU91160 standard; Protein; 257 AA.
XX
AC   AAU91160;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Synechocystis pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Synechocystis sp.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 77-78; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   257 AA;

  Query Match             10.7%;  Score 116.5;  DB 23;  Length 257;
  Best Local Similarity   30.4%;  Pred. No. 0.00025;
  Matches   52;  Conservative   26;  Mismatches   66;  Indels   27;  Gaps   11;

Qy          49 LKNQKNFINLEPYFLFDTIYQGLGIDRIAA--CYTIEDGVVVDAGSAITIDIISNSIHL- 105
               |||    :|| | |  |    |||        |      :||| |:|:||        |
Db          87 LKNLP-LVNLYPSFGIDRALAGLGTGLTYGFPC------LVVDGGTALTITGFDQDKKLV 139

Qy         106 GGFILPGIA-NYKKIYSHIS--PRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKD 162
               || ||||:      :   ::  |:|: :  |::  | :   |  |:  ||  |:   ::
Db         140 GGAILPGLGLQLATLGDRLAALPKLEMDQLTELP-DRWALDTPSAIFSGVVYGVLGALQS 198

Qy         163 AAQN-KKLY------FTGGDGQFLANYF-DHA-----IYDKLLIFRGMKKI 200
                 |: :||:       |||||: |  :  :|:      :|  ||| ||  |
Db         199 YLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNLIFLGMAAI 249
```

```
RESULT 12
AAU91178
ID   AAU91178 standard; Protein; 249 AA.
XX
AC   AAU91178;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #16.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas putida.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54199.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 108-109; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match             10.5%;  Score 114.5;  DB 23;  Length 249;
  Best Local Similarity   26.6%;  Pred. No. 0.00039;
  Matches   42;  Conservative   28;  Mismatches   49;  Indels   39;  Gaps    8;

Qy         68 YQGLGIDR-IAA--CYTIEDG--VVVDAGSAITIDIIS-NSIHLGGFILPGIANYKKIYS 121
              || ||:|| :||   : :  |  :|:| |:|   | :| :  ||||:| ||:
Db         96 YQRLGMDRWLAALGAFHLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGM-------- 147

Qy        122 HISPRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK---- 167
                 | ::|:  |        ||  :: : :|| |      | :|  |:::  |  :
Db        148 ---PLMRSQLRTHTRRIRYDDASAERALSSLSPGRSTVEAVERGCVLMLQGFAYTQLEQA 204

Qy        168 --------KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                     :: ||||   :      |     |:| |:
Db        205 RVLWGEEFTVFLTGGDAPLVRAALPQARVVPDLVFVGL 242


RESULT 13
AAU91182
ID   AAU91182 standard; Protein; 249 AA.
XX
AC   AAU91182;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #20.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas syringae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54203.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 113-114; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match             10.4%;  Score 113.5;  DB 23;  Length 249;
  Best Local Similarity   26.5%;  Pred. No. 0.00051;
  Matches   41;  Conservative   26;  Mismatches   49;  Indels   39;  Gaps    7;

Qy         71 LGIDR---IAACYTIEDG--VVVDAGSAITIDII-SNSIHLGGFILPGIANYKKIYSHIS 124
              ||:||       | :  |  :|:| |:||| | : ::  |||||| ||:
Db         99 LGLDRWLAFVGAYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGM---------- 147

Qy        125 PRLKSEFNTQVSL----DAFPQKTMDALSYG------VFKGIYLLIKDAAQNK------- 167
               | ::::  |        ||  :| :  |  |      | :|  |:::  |  :
Db        148 PLMRNQLRTHTRRIRYDDAEAEKALVRLVPGRATAEAVERGCSLMLRGFAMTQIEIAREY 207

Qy        168 -----KLYFTGGDGQFLANYFDHAIYDKLLIFRGM 197
                   :: ||||   :|:    |     |:| |:
Db        208 WGDDFAIFVTGGDAVLVADVLPGARIVPDLVFVGL 242


RESULT 14
AAU91170
ID   AAU91170 standard; Protein; 262 AA.
XX
AC   AAU91170;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #8.
XX
```

```
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus anthracis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54191.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 8; Page 98-99; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   262 AA;

  Query Match             10.3%;  Score 113;  DB 23;  Length 262;
  Best Local Similarity   21.8%;  Pred. No. 0.00062;
  Matches   57;  Conservative   50;  Mismatches   93;  Indels   62;  Gaps   13;

Qy          1 MLLCDIGNSNANF------------LDDNKYFT-----LNIDQFLE-----FKNEQKIF 37
              : : |:||:||                :: ::: |     : : | ||     |:: : |
Db          2 IFVLDVGNTNAVLGVFEEGELRQHWRMETDRHKTEDEYGMLVKQLLEHEGLSFEDVKGII 61

Qy         38 YINVNEHLKEHLKNQ-KNFINLEPYFLFDTIYQGL----------GIDRIAACYTIEDG- 85
                :|   :   |:   : :  ::|  :   |  ||          | |||        |
Db         62 VSSVVPPIMFALERMCEKYFKIKPLVVGPGIKTGLNIKYENPREVGADRIVNAVA---GI 118

Qy         86 -------VVVDAGSAITIDIISNSIH-LGGFILPGI-ANYKKIYSHISPRLKSEFNTQVS 136
                    ::|| |:| |   |:   | :|| | |||  : : :||  :   : |     |
Db        119 HLYGSPLIIVDFGTATTYCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRIEITKPSS 178

Qy        137 LDAFPQKTMDALSYGVFKG--------IYLLIKDAAQNKKLYFTGGDGQFL---ANYFDH 185
              :    : |: |:  |:  |        :  : ::| |  |:  |||  : :   :|  |
Db        179 V--VGKNTVSAMQSGILYGYVGQVEGIVKRMKEEAKQEPKVIATGGLAKLISEESNVID- 235

Qy        186 AIYDKLLIFRGMKKIIKENPNL 207
               : |  |  :|:  : : | ||
Db        236 -VVDPFLTLKGLYMLYERNANL 256


RESULT 15
AAU91173
ID   AAU91173 standard; Protein; 260 AA.
XX
AC   AAU91173;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #11.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Caulobacter crescentus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54194.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 102-103; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   260 AA;

  Query Match             10.2%;  Score 111;  DB 23;  Length 260;
  Best Local Similarity   26.8%;  Pred. No. 0.001;
  Matches   52;  Conservative   31;  Mismatches   64;  Indels   47;  Gaps   11;

Qy         48 HLKN-QKNFINLEPYFLFDTIYQGL----------GIDRI-----AACYTIEDGVVVDAG 91
              :|:|  : : |:||  : :    |:          | ||:     ||       ||:|:|
Db         72 NLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSG 131

Qy         92 SAITIDII-SNSIHLGGFILPGIANYKKIYSHIS----PRL---KSEFNTQVSLDAFPQK 143
              :| | ||: ::    || | ||| |      | :    ||:   :   |  |  |
Db        132 TATTFDIVAADGAFEGGIIAPGI-NLSMQALHEAAAKLPRIAIQRPAGNRIVGTD----- 185

Qy        144 TMDALSYGVFKGIYLLIK--------DAAQNKKLYFTGGDGQFLANYFDHAI-----YDK 190
              |: |:  ||| |   ||:        :  :   :  |||    :|: |: |      :|
Db        186 TVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGG----VASLFEGATDSIDHFDS 241

Qy        191 LLIFRGMKKIIKEN 204
               |  ||: :| : |
Db        242 DLTIRGLLEIYRRN 255


Search completed: June 24, 2003, 21:46:25
Job time : 23.4576 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 10.1672 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-14
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

otal number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	9	US-09-895-913A-74	Sequence 74, Appl
2	1153	100.0	223	9	US-09-813-453A-14	Sequence 14, Appl
3	1153	100.0	223	9	US-09-813-453A-67	Sequence 67, Appl
4	391.5	34.0	209	9	US-09-813-453A-21	Sequence 21, Appl
5	165	14.3	246	9	US-09-813-453A-9	Sequence 9, Appli
6	159.5	13.8	248	9	US-09-813-453A-20	Sequence 20, Appl
7	148	12.8	249	9	US-09-813-453A-70	Sequence 70, Appl
8	144.5	12.5	241	9	US-09-813-453A-63	Sequence 63, Appl
9	144.5	12.5	249	9	US-09-813-453A-61	Sequence 61, Appl
10	141.5	12.3	257	9	US-09-813-453A-13	Sequence 13, Appl
11	138.5	12.0	229	9	US-09-813-453A-12	Sequence 12, Appl
12	129.5	11.2	244	9	US-09-813-453A-41	Sequence 41, Appl
13	116.5	10.1	592	9	US-09-813-453A-22	Sequence 22, Appl
14	116.5	10.1	592	9	US-09-813-453A-43	Sequence 43, Appl
15	116	10.1	460	9	US-09-813-453A-39	Sequence 39, Appl
16	115.5	10.0	242	9	US-09-813-453A-65	Sequence 65, Appl
17	109.5	9.5	233	9	US-09-813-453A-17	Sequence 17, Appl
18	109.5	9.5	258	9	US-09-813-453A-2	Sequence 2, Appli
19	108	9.4	219	9	US-09-813-453A-57	Sequence 57, Appl
20	105.5	9.2	257	9	US-09-813-453A-53	Sequence 53, Appl
21	102.5	8.9	250	9	US-09-813-453A-3	Sequence 3, Appli
22	102	8.8	212	9	US-09-813-453A-59	Sequence 59, Appl
23	101	8.8	256	9	US-09-813-453A-55	Sequence 55, Appl
24	100.5	8.7	255	9	US-09-813-453A-7	Sequence 7, Appli
25	98.5	8.5	262	9	US-09-813-453A-45	Sequence 45, Appl
26	91	7.9	254	9	US-09-813-453A-47	Sequence 47, Appl
27	89.5	7.8	272	9	US-09-712-363-276	Sequence 276, App
28	89.5	7.8	272	9	US-09-813-453A-5	Sequence 5, Appli
29	88.5	7.7	262	9	US-09-813-453A-11	Sequence 11, Appl
30	87.5	7.6	258	9	US-09-813-453A-49	Sequence 49, Appl
31	87	7.5	947	10	US-09-801-574-4	Sequence 4, Appli
32	84	7.3	764	10	US-09-815-242-11556	Sequence 11556, A
33	82.5	7.2	265	9	US-09-813-453A-4	Sequence 4, Appli
34	82	7.1	688	10	US-09-815-242-11416	Sequence 11416, A
35	81.5	7.1	273	9	US-09-813-453A-10	Sequence 10, Appl
36	80	6.9	260	9	US-09-813-453A-51	Sequence 51, Appl
37	79	6.9	310	10	US-09-815-242-10283	Sequence 10283, A
38	79	6.9	375	9	US-09-981-353-113	Sequence 113, App
39	79	6.9	2799	9	US-10-151-736-4	Sequence 4, Appli
40	78	6.8	770	9	US-10-153-668-7	Sequence 7, Appli
41	77.5	6.7	258	9	US-09-813-453A-6	Sequence 6, Appli
42	77	6.7	692	9	US-10-101-464A-897	Sequence 897, App
43	76	6.6	434	9	US-10-153-668-454	Sequence 454, App
44	76	6.6	439	9	US-09-374-046A-44	Sequence 44, Appl
45	76	6.6	653	9	US-10-174-590-402	Sequence 402, App

ALIGNMENTS

RESULT 1
US-09-895-913A-74
; Sequence 74, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1el Helicobacter Polypeptides in
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-74

Query Match 100.0%; Score 1153; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e-107;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
```

```
Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 2
US-09-813-453A-14
; Sequence 14, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-813-453A-14

```
  Query Match             100.0%;  Score 1153;  DB 9;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-107;
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 3
US-09-813-453A-67
; Sequence 67, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-813-453A-67

```
  Query Match             100.0%;  Score 1153;  DB 9;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-107;
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 4
US-09-813-453A-21
; Sequence 21, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
US-09-813-453A-21

```
  Query Match             34.0%;  Score 391.5;  DB 9;  Length 209;
  Best Local Similarity   39.2%;  Pred. No. 2.3e-31;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::|||||:  :|  : : |:   :       :::||||:|||  :: | |   | |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
```

RESULT 5
US-09-813-453A-9
; Sequence 9, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers

```
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Thermotoga maritima
US-09-813-453A-9

  Query Match             14.3%;  Score 165;  DB 9;  Length 246;
  Best Local Similarity   23.2%;  Pred. No. 1.3e-08;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
                 :| |:|||                |:           |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 ----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                        :: |:||   |  :|       |   | || :                :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|       ||:::| |:| |:||:  | : || ||||    :|: :: :|||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

Qy        144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
                |  |   |: | | : :  |:|         :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
                | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 6
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

  Query Match             13.8%;  Score 159.5;  DB 9;  Length 248;
  Best Local Similarity   28.6%;  Pred. No. 4.6e-08;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
                 :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::

Db         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                    :|:| |:|:| ||:  :| |||| | ||:              :|  |  : |
Db        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
                 || :|          | :||  | :| :         :|| : |  : :|:| ||||:
Db        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy        191 YLSAFLPHSVCKERLVFDGMEIA 213
                :   |  :     ||| |: :|
Db        222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 7
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             12.8%;  Score 148;  DB 9;  Length 249;
  Best Local Similarity   29.3%;  Pred. No. 6.6e-07;
  Matches   58;  Conservative   34;  Mismatches   78;  Indels   28;  Gaps    7;

Qy         34 SAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACL 93
                 ||:|  | :    : | ||           |   |  :|| :   |:  ||:|| :| :
Db         58 SAEETAKLVCALADTFSISP---------VCAAPAPELAGVINGYDDFARLGLDRWLAFV 108

Qy         94 AVNNGV-----VVDAGSAITIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILE----QP 143
                 : |      |:| |:||| | ::  |||||| | ||:    :  :   : :      :
Db        109 GAYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEA 168

Qy        144 FKALDSLEVLPKSTRDAVNYG--MVLSVIACIQ------HLAKNQKIYLCGGDAKYLSAF 195
                 |||  | |  ::| :|:|  |  ::|   |  |       :   :  |:: |||| ::
Db        169 EKALVRL-VPGRATAEAVERGCSLMLRGFAMTQIEIAREYWGDDFAIFVTGGDAVLVADV 227

Qy        196 LPHSVCKERLVFDGMEIA 213
                || :     ||| |: :|
Db        228 LPGARIVPDLVFVGLALA 245


RESULT 8
US-09-813-453A-63
; Sequence 63, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
```

```
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
;   LENGTH: 241
;   TYPE: PRT
;   ORGANISM: Thiobacillus ferrooxidans
US-09-813-453A-63

  Query Match             12.5%;  Score 144.5;  DB 9;  Length 241;
  Best Local Similarity   24.4%;  Pred. No. 1.4e-06;
  Matches   60;  Conservative   40;  Mismatches   97;  Indels   49;  Gaps   10;

Qy         13 LVLCDIGNTRIHFAQNYQ--LFSSAK--------EDLKRLGIQKEIFYISVNEENEKALL 62
              ::   :||||   |  :    | |          | |:: |:     ::|       ||
Db          1 MIFIAVGNTRTLLAHTHDGVHFDSVSVATSLPPTEILQQPGLT----WLSAPNREPVALG 56

Qy         63 NCYPNA--------------KNIAGFFH--LETDY---VGLGIDRQMACLAV------NN 97
                | |               :   |||   :  ||     || ||:   ||       :
Db         57 GVVPAALAAWREALATAEVREPDPGFFRRAVPHDYHPPESLGFDRRCCLLAAAMDYPGQD 116

Qy         98 GVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDS-LEVLPKS 156
               :|:| |:||||||:  |   || ||||:|  :    :  :|  |   |::  |:|
Db        117 SIVIDMGTAITIDLLAGGHFRGGRILPGIAMSLRGLHEGTALL--PEVVLNAPAEMLGND 174

Qy        157 TRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDG 209
              | :|:  |::          |   :  :   :| : ||||:     :  |: :  |:  |
Db        175 TSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRG 234

Qy        210 MEIALK 215
                : ::
Db        235 FYLWIR 240


RESULT 9
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
   PRIOR APPLICATION NUMBER: US 60/227,860
   PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             12.5%;  Score 144.5;  DB 9;  Length 249;
  Best Local Similarity   29.8%;  Pred. No. 1.5e-06;
  Matches   61;  Conservative   26;  Mismatches   63;  Indels   55;  Gaps   10;

Qy         51 ISVNEENEKALLNC-------------YPNAKNIAGFFHLETDYVGLGIDRQMAC---- 92
              :||  | |   | |              :|  : :||  :   ||  ||:|| :|
Db         54 VSVRSEEETDAL-CALIAQAFAVQAKVAHP-VREMAGVRNGYDDYQRLGMDRWLAALGAF 111

Qy         93 -LAVNNGVVVDAGSAITIDLIK-EGKHLGGCILPGL----------AQYIHAYKKSAKIL 140
               ||    :|:| |:|   | :  :|:|||| | ||:           : |     ||:
Db        112 HLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRIRYDDASAE-- 169

Qy        141 EQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ------------KIYLCGGD 188
                :|| ||    :|| :||  | ||     :|  |  |              ::| |||
Db        170 ----RALSSLSP-GRSTVEAVERGCVL----MLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        189 AKYLSAFLPHSVCKERLVFDGMEIA 213
                |  : | || :     ||| |: :|
Db        221 APLVRAALPQARVVPDLVFVGLAMA 245


RESULT 10
US-09-813-453A-13
; Sequence 13, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Synechocystis sp.
US-09-813-453A-13

  Query Match             12.3%;  Score 141.5;  DB 9;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 3.1e-06;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy         18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
              |||:|:|:|     : |    |:                ||    :    | |: |
Db         25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy         59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                         |:| ||:                ||||  :| |           :|||
Db         80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy        104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
              |:|:||    : | | || |||||   :       ::   |   :| |  ||      |
Db        124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy        159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
               |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||        : |
Db        182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

Qy        206 VFDGM 210
               :| ||
Db        242 IFLGM 246


RESULT 11
US-09-813-453A-12
; Sequence 12, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
```

```
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Aquifex aeolicus
US-09-813-453A-12

  Query Match             12.0%;  Score 138.5;  DB 9;  Length 229;
  Best Local Similarity   26.5%;  Pred. No. 5.2e-06;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy         17 DIGNTRIHFA------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
              |:||: :  |              : :       || ||       ||| :   : :
Db          7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy         64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                 |  | :    | :: ||     || ||     :       | ||: ||:|: |||: |
Db         60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy        115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
              ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
Db        120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172

Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                 : | :   : : | : |: : ||: || | |    :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219


RESULT 12
US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 244
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

  Query Match             11.2%;  Score 129.5;  DB 9;  Length 244;
  Best Local Similarity   26.0%;  Pred. No. 4.5e-05;
  Matches   58;  Conservative   40;  Mismatches   82;  Indels   43;  Gaps   12;

Qy         12 NLVLCDIGNT--RIHFAQNYQLFS------SAKEDLKRL----GIQKEIF-YISVNEENE 58
              ||:: | ||:  :: | :|  : |       | : |  |       | |:  : : :|
Db          4 NLIV-DQGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEA 62

Qy         59 KALLNCYPNAKNIAGF-----FHLETDYVGLGIDRQMACLAVNN------GVVVDAGSAI 107
              :|::     |  : |       |: |   || ||  | :  ::       :|:|||:||
Db         63 EAIVRSCAAASLMMGTETPVPLRLQYDRRTLGADRLAAVVGAHSLYPNTELLVIDAGTAI 122

Qy        108 TIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK------STRDA 160
              | : :  || :||| | |||    |   |:  :      :|  : ||      || :|
Db        123 TYERVSAEGIYLGGNISPGL----HLRFKALHLFTGRLPLIDPSGISPKIAEYGSSTEEA 178

Qy        161 VNYGMVLSVIACIQ------HLAKNQK-IYLCGGDAKYLSAFL 196
              :  |:: :  |       |  : :  : | |||| ||:  :
Db        179 ITAGVIHGLAGEIDRYIDDLHAKEGRSAVILTGGDANYLARII 221


RESULT 13
US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-22

  Query Match             10.1%;  Score 116.5;  DB 9;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0029;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                 : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||          | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 14
US-09-813-453A-43
; Sequence 43, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-43
```

```
  Query Match             10.1%;  Score 116.5;  DB 9;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0029;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |       | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   | |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 15
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
S-09-813-453A-39

  Query Match             10.1%;  Score 116;  DB 9;  Length 460;
  Best Local Similarity   25.5%;  Pred. No. 0.0023;
  Matches   62;  Conservative   37;  Mismatches   88;  Indels   56;  Gaps   12;

Qy          9 DLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KEIFYISVNE 55
              | :  :| : ||:|: :|  :|     : |:   ||  || :          |   :|
Db        204 DSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCG 263

Qy         56 ENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-----AVNNG 98
              |::||          :   |::    |    : | |       | ||    |     : |
Db        264 ESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGSRRFSRNAC 319

Qy         99 VVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI----LEQPFKALDSLEVL 153
              |||  |:|:|:| |  :| :||| |:||     |  |:|  :    | :|
Db        320 VVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRP---AGKRYPF 372

Qy        154 PKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDAKYLSAFLPHSVCKE 203
              | :| :||  ||: :|   |  :           |   : : || |  ::  || :   |
Db        373 PTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAE 432

Qy        204 RLV 206
              |
Db        433 NTV 435
```

```
Search completed: June 24, 2003, 22:24:02
Job time : 11.1672 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 6.51452 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-14
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

otal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	89.5	7.8	1582	2	US-08-404-531B-9	Sequence 9, Appli
2	89.5	7.8	1582	3	US-08-476-900A-9	Sequence 9, Appli
3	89.5	7.8	1582	3	US-08-488-546A-9	Sequence 9, Appli
4	89.5	7.8	1582	4	US-08-726-320-5	Sequence 5, Appli
5	89.5	7.8	1582	4	US-09-208-716-5	Sequence 5, Appli
6	86.5	7.5	1581	2	US-08-404-531B-6	Sequence 6, Appli
7	86.5	7.5	1581	3	US-08-476-900A-6	Sequence 6, Appli
8	86.5	7.5	1581	3	US-08-488-546A-6	Sequence 6, Appli
9	86.5	7.5	1581	4	US-08-726-320-4	Sequence 4, Appli
10	86.5	7.5	1581	4	US-09-208-716-4	Sequence 4, Appli
11	81	7.0	1580	4	US-08-726-320-1	Sequence 1, Appli
12	81	7.0	1580	4	US-09-208-716-1	Sequence 1, Appli
13	81	7.0	1581	4	US-08-726-320-3	Sequence 3, Appli
14	81	7.0	1581	4	US-09-208-716-3	Sequence 3, Appli
15	79.5	6.9	424	6	5169835-6	Patent No. 5169835
16	78	6.8	382	4	US-09-134-001C-5449	Sequence 5449, Ap
17	77.5	6.7	1447	4	US-09-376-330-17	Sequence 17, Appl
18	77	6.7	601	4	US-09-134-001C-3791	Sequence 3791, Ap
19	76	6.6	374	2	US-07-857-224B-82	Sequence 82, Appl
20	76	6.6	374	2	US-07-857-224B-83	Sequence 83, Appl
21	75.5	6.5	375	4	US-09-347-878-56	Sequence 56, Appl
22	74.5	6.5	248	4	US-09-081-689-4	Sequence 4, Appli
23	74.5	6.5	248	4	US-08-858-207A-340	Sequence 340, App
24	74.5	6.5	442	4	US-09-081-689-2	Sequence 2, Appli
25	74.5	6.5	442	4	US-09-305-984-14	Sequence 14, Appl
26	74.5	6.5	442	4	US-09-073-541A-14	Sequence 14, Appl
27	74	6.4	513	4	US-09-134-001C-4490	Sequence 4490, Ap
28	73.5	6.4	904	6	5244792-4	Patent No. 5244792
29	72.5	6.3	903	1	US-08-220-151-8	Sequence 8, Appli
30	72.5	6.3	903	1	US-08-413-118-8	Sequence 8, Appli
31	72.5	6.3	903	3	US-08-473-446-8	Sequence 8, Appli
32	72.5	6.3	1012	2	US-08-680-326-34	Sequence 34, Appl
33	72.5	6.3	1462	3	US-07-792-600-31	Sequence 31, Appl
34	72.5	6.3	1462	3	US-09-157-021-31	Sequence 31, Appl
35	72.5	6.3	1462	3	US-09-156-842-31	Sequence 31, Appl
36	72	6.2	248	6	5169835-15	Patent No. 5169835
37	71.5	6.2	307	4	US-08-804-166-4	Sequence 4, Appli
38	71.5	6.2	307	4	US-08-910-991-4	Sequence 4, Appli
39	70.5	6.1	511	2	US-09-019-201A-2	Sequence 2, Appli
40	70.5	6.1	579	1	US-08-126-564A-31	Sequence 31, Appl
41	70.5	6.1	579	5	PCT-US94-09143-31	Sequence 31, Appl
42	70	6.1	355	1	US-08-008-688A-2	Sequence 2, Appli
43	70	6.1	613	2	US-08-915-207-2	Sequence 2, Appli
44	70	6.1	613	2	US-08-915-207-4	Sequence 4, Appli
45	70	6.1	613	4	US-09-238-555-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-404-531B-9
; Sequence 9, Application US/08404531B
; Patent No. 5863724
; GENERAL INFORMATION:
; APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
; APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
; TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 5863724
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
; ADDRESSEE: No. 5863724ris
; STREET: One Liberty Place 46th. Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/404,531B
; FILING DATE: 15-MAR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Beardell, Lori Y.
; REGISTRATION NUMBER: 34,293
; REFERENCE/DOCKET NUMBER: BYLR-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-404-531B-9

Query Match 7.8%; Score 89.5; DB 2; Length 1582;
Best Local Similarity 22.7%; Pred. No. 0.83;
Matches 58; Conservative 44; Mismatches 81; Indels 73; Gaps 15;

```
Qy          14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
               |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db        1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223
```

```
Qy          65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                   | :: |||  |        ||   :|:|  :     |  :::|         |:|
Db        1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy         103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
                 |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db        1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337

Qy         156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
               :  |          :| |   |  :  ::|    ::  ||| :||  |  |     |: |
Db        1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy         198 HSVCKERLVFDGMEIA 213
                 : : |:: ||::||
Db        1396 VDMFEGRIIIDGIDIA 1411


RESULT 2
US-08-476-900A-9
; Sequence 9, Application US/08476900A
; Patent No. 6031150
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6031150
;    TITLE OF INVENTION:  and Method of Detecting Persistent Hyperinsulinemic Hypoglycem
;    TITLE OF INVENTION:  Infancy
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 6031150ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,900A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1582 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-476-900A-9

  Query Match             7.8%;  Score 89.5;  DB 3;  Length 1582;
  Best Local Similarity   22.7%;  Pred. No. 0.83;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy          14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
               |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db        1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy          65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                   | :: |||  |        ||   :|:|  :     |  :::|         |:|
Db        1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy         103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
                 |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db        1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337

Qy         156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
               :  |          :| |   |  :  ::|    ::  ||| :||  |  |     |: |
Db        1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy         198 HSVCKERLVFDGMEIA 213
                 : : |:: ||::||
Db        1396 VDMFEGRIIIDGIDIA 1411


RESULT 3
US-08-488-546A-9
; Sequence 9, Application US/08488546A
; Patent No. 6054313
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT:  Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6054313
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 6054313ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,546A
;      FILING DATE:  07-JUNE-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/404,531
;      FILING DATE:  15-MARCH-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1582 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-488-546A-9

  Query Match             7.8%;  Score 89.5;  DB 3;  Length 1582;
  Best Local Similarity   22.7%;  Pred. No. 0.83;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy          14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
               |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db        1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy          65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                   | :: |||  |        ||   :|:|  :     |  :::|         |:|
Db        1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy         103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
                 |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db        1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337
```

```
Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |         :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
               : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411
```

RESULT 4
US-08-726-320-5
; Sequence 5, Application US/08726320
; Patent No. 6171815
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Coleman, Roger
; TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,320
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: AF-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 784874
US-08-726-320-5

Query Match 7.8%; Score 89.5; DB 4; Length 1582;
Best Local Similarity 22.7%; Pred. No. 0.83;
Matches 58; Conservative 44; Mismatches 81; Indels 73; Gaps 15;

```
Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337
```

```
Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |         :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
               : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411
```

RESULT 5
US-09-208-716-5
; Sequence 5, Application US/09208716
; Patent No. 6235880
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Coleman, Roger
; TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,716
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/726,320
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: AF-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 784874
US-09-208-716-5

Query Match 7.8%; Score 89.5; DB 4; Length 1582;
Best Local Similarity 22.7%; Pred. No. 0.83;
Matches 58; Conservative 44; Mismatches 81; Indels 73; Gaps 15;

```
Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
```

```
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 6
US-08-404-531B-6
; Sequence 6, Application US/08404531B
; Patent No. 5863724
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT:  Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 5863724
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5863724ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/404,531B
;      FILING DATE:  15-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
S-08-404-531B-6

  Query Match             7.5%;  Score 86.5;  DB 2;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 7
US-08-476-900A-6
; Sequence 6, Application US/08476900A
; Patent No. 6031150
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6031150
;    TITLE OF INVENTION:  and Method of Detecting Persistent Hyperinsulinemic Hypogl
;    TITLE OF INVENTION:  Infancy
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 6031150ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,900A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-476-900A-6

  Query Match             7.5%;  Score 86.5;  DB 3;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 8
US-08-488-546A-6
; Sequence 6, Application US/08488546A
```

```
; Patent No. 6054313
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT:  Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor          a
; Patent No. 6054313
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 6054313ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,546A
;      FILING DATE:  07-JUNE-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/404,531
;      FILING DATE:  15-MARCH-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-488-546A-6

  Query Match             7.5%;  Score 86.5;  DB 3;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |          :  :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : ||   |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |          :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 9
US-08-726-320-4
; Sequence 4, Application US/08726320
; Patent No. 6171815
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,320
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  Genbank
;      CLONE:  1311534
US-08-726-320-4

  Query Match             7.5%;  Score 86.5;  DB 4;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |          :  :|:|  :    |  :::|         |:|
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1280

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : ||   |:     : :: | :  ::||
Db       1281 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1336

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |          :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1337 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1394

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1395 VDMFEGRIIIDGIDIA 1410


RESULT 10
US-09-208-716-4
; Sequence 4, Application US/09208716
; Patent No. 6235880
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
```

```
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,716
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,320
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  Genbank
;      CLONE:  1311534
US-09-208-716-4

  Query Match             7.5%;  Score 86.5;  DB 4;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |          :  :|:|  :    |  :::|          |:|
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1280

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:         ::|    |  |  : ||   |:     : :: | :  ::||
Db       1281 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1336

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              : |         :| |  |  :  ::|    ::  ||| :||  |  |    |: |
Db       1337 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1394

Qy        198 HSVCKERLVFDGMEIA 213
               : : |:: ||::||
Db       1395 VDMFEGRIIIDGIDIA 1410


RESULT 11
US-08-726-320-1
; Sequence 1, Application US/08726320
; Patent No. 6171815
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,320
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1580 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-08-726-320-1

  Query Match             7.0%;  Score 81;  DB 4;  Length 1580;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC---------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                   | ::
Db       1174 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1231

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |          :  :|:|  :    |  :::|         |:|   |  :|  |:
Db       1232 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1288

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                       ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1289 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1345

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |       :  ::
Db       1346 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1403

Qy        208 DGMEIA 213
               ||::||
Db       1404 DGIDIA 1409


RESULT 12
US-09-208-716-1
; Sequence 1, Application US/09208716
; Patent No. 6235880
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
```

```
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,716
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,320
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1580 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-09-208-716-1

  Query Match             7.0%;  Score 81;  DB 4;  Length 1580;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                 | ::
Db       1174 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1231

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |         :  :|:|  :    |  :::|        |:|   |  :|  |:
Db       1232 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1288

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                  ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1289 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1345

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1346 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1403

Qy        208 DGMEIA 213
               ||::||
Db       1404 DGIDIA 1409


RESULT 13
US-08-726-320-3
; Sequence 3, Application US/08726320
; Patent No. 6171815
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,320
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1369844
US-08-726-320-3

  Query Match             7.0%;  Score 81;  DB 4;  Length 1581;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                 | ::
Db       1175 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1232

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |         :  :|:|  :    |  :::|        |:|   |  :|  |:
Db       1233 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1289

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                  ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1290 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1346

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1347 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1404

Qy        208 DGMEIA 213
               ||::||
Db       1405 DGIDIA 1410


RESULT 14
US-09-208-716-3
; Sequence 3, Application US/09208716
; Patent No. 6235880
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;       STREET:  3174 Porter Drive
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  U.S.
;       ZIP:  94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/208,716
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/726,320
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Billings, Lucy J.
;       REGISTRATION NUMBER:  36,749
;       REFERENCE/DOCKET NUMBER:  AF-0001 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  415-855-0555
;       TELEFAX:  415-845-4166
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1581 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     IMMEDIATE SOURCE:
;       LIBRARY:  GenBank
;       CLONE:  1369844
US-09-208-716-3

  Query Match             7.0%;  Score 81;  DB 4;  Length 1581;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                 | ::
Db       1175 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1232

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |         :  :|:|  :    |  :::|         |:|   |  :|  |:
Db       1233 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1289

         113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                  ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1290 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1346

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1347 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1404

Qy        208 DGMEIA 213
               ||::||
Db       1405 DGIDIA 1410


RESULT 15
5169835-6
;Patent No. 5169835
;     APPLICANT: WAI-YEE, CHAN
;     TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS
;     NUMBER OF SEQUENCES: 48
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/390,409
;       FILING DATE: 07-AUG-1989
;SEQ ID NO:6:
;       LENGTH: 424
```

```
5169835-6

  Query Match             6.9%;  Score 79.5;  DB 6;  Length 424;
  Best Local Similarity   22.3%;  Pred. No. 1.5;
  Matches   40;  Conservative   23;  Mismatches   61;  Indels   55;  Gaps    7;

Qy         56 ENEKALLNCYPNAKNIAGF----------FHLETDYVGLGIDRQM-------------- 90
              | :  ||  :   :|: |:          :|  | ||   :| |:
Db         47 EGKDVLLLVHNLPQNLTGYIWYKGQMTDLYHYITSYV---VDGQIIYGPAYSGRETVYSN 103

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
              | |.: |    ||||  |: :|| |   ||         :  |        | |  :: |
Db        104 ASLLIQNVTQEDAGS-YTLHIIKRGDGTGG---------VTGYFTVTLYSETPKPSISSS 153

Qy        151 EVLPKSTRDAVNY----------------GMVLSVIACIQHLAKNQKIYLCGGDAKYLS 193
               : |:   :||                 |  | :   :|    |: :||  |  ||::
Db        154 NLNPREVMEAVRLICDPETPDASYLWLLNGQNLPMTHRLQLSKTNRTLYLF-GVTKYIA 211


Search completed: June 24, 2003, 21:49:29
Job time : 7.51452 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 18.1879 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-14
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	16	O25533	O25533 helicobacte
2	1115	96.7	223	16	Q9ZKY6	Q9zky6 helicobacte
3	391.5	34.0	209	16	Q9PIA9	Q9pia9 campylobact
4	165	14.3	246	16	Q9WZY5	Q9wzy5 thermotoga
5	159.5	13.8	248	16	Q9HWC1	Q9hwc1 pseudomonas
6	158	13.7	276	16	Q8YQD7	Q8yqd7 anabaena sp
7	141.5	12.3	257	16	P74045	P74045 synechocyst
8	139	12.1	256	16	Q8RFE4	Q8rfe4 fusobacteri
9	138.5	12.0	229	16	O67753	O67753 aquifex aeo
10	126.5	11.0	259	16	Q8XHL5	Q8xhl5 clostridium
11	117.5	10.2	273	16	Q97EB4	Q97eb4 clostridium
12	116.5	10.1	592	16	Q9JXF1	Q9jxf1 neisseria m
13	116.5	10.1	592	16	Q9JWI7	Q9jwi7 neisseria m
14	116	10.1	224	16	Q98Q93	Q98q93 mycoplasma
15	115.5	10.0	242	16	Q9PCI4	Q9pci4 xylella fas
16	102	8.8	212	2	O32514	O32514 desulfovibr
17	100.5	8.7	295	16	Q8Y2M4	Q8y2m4 ralstonia s
18	94.5	8.2	255	16	Q8R7M2	Q8r7m2 thermoanaer
19	94.5	8.2	338	5	O76462	O76462 drosophila
20	94	8.2	407	16	O31557	O31557 bacillus su
21	93.5	8.1	1164	10	Q9XIK6	Q9xik6 arabidopsis
22	91.5	7.9	440	17	Q8ZZ98	Q8zz98 pyrobaculum
23	91.5	7.9	683	11	Q9Z2P1	Q9z2p1 mus musculu
24	91	7.9	254	16	Q9KGH5	Q9kgh5 bacillus ha
25	90.5	7.8	274	16	Q9CD56	Q9cd56 mycobacteri
26	90.5	7.8	482	16	Q9X0V6	Q9x0v6 thermotoga
27	90	7.8	313	16	Q8XKL3	Q8xkl3 clostridium
28	89.5	7.8	272	16	O06282	O06282 mycobacteri
29	89	7.7	243	12	Q98237	Q98237 molluscum c
30	89	7.7	594	5	Q9VVK8	Q9vvk8 drosophila
31	89	7.7	1483	10	Q8S1U8	Q8slu8 oryza sativ
32	88.5	7.7	258	2	Q9F985	Q9f985 bacillus st
33	88.5	7.7	262	16	O51477	O51477 borrelia bu
34	88.5	7.7	493	12	Q9PYQ7	Q9pyq7 xestia c-ni
35	88	7.6	274	16	Q8RI50	Q8ri50 fusobacteri
36	87.5	7.6	383	2	Q9FD71	Q9fd71 enterococcu
37	87.5	7.6	483	8	Q35316	Q35316 oryza sativ
38	87	7.5	947	11	Q99MW9	Q99mw9 mus musculu
39	87	7.5	1051	17	Q8TMS3	Q8tms3 methanosarc
40	86.5	7.5	393	16	Q8YRY5	Q8yry5 anabaena sp
41	86	7.5	474	16	Q9PQV7	Q9pqv7 ureaplasma
42	86	7.5	702	16	Q8UD86	Q8ud86 agrobacteri
43	85.5	7.4	240	16	Q8ZKR0	Q8zkr0 salmonella
44	85.5	7.4	240	16	Q8Z2X1	Q8z2x1 salmonella
45	85.5	7.4	401	16	Q9KFC0	Q9kfc0 bacillus ha

ALIGNMENTS

RESULT 1
O25533
ID O25533 PRELIMINARY; PRT; 223 AA.
AC O25533;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein HP0862.
GN HP0862.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL; AE000596; AAD07916.1; -.
DR TIGR; HP0862; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 223 AA; 24655 MW; C1D13E8BEC1EB9AF CRC64;

Query Match 100.0%; Score 1153; DB 16; Length 223;
Best Local Similarity 100.0%; Pred. No. 2.2e-94;

```
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223


RESULT 2
 9ZKY6
 D   Q9ZKY6      PRELIMINARY;      PRT;   223 AA.
AC   Q9ZKY6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PUTATIVE.
GN   JHP0796.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001509; AAD06372.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   223 AA;  24754 MW;  B31377DA1AEF6E0E CRC64;

  Query Match             96.7%;  Score 1115;  DB 16;  Length 223;
  Best Local Similarity   94.2%;  Pred. No. 5e-91;
  Matches  210;  Conservative   11;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||| |||:|:||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFKDLKDLILCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||:|||||||||||| |||:|||||||||||:||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYIGLGIDRQMACLAVVNGVIVDAGSAITIDLVKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              |||||||||:||||||||||||||||||||||:|||||||||:||:|:|||||||:|
Db        121 CILPGLAQYVHAYKKSAKILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCIQHLAKDQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223


RESULT 3
Q9PIA9
ID   Q9PIA9      PRELIMINARY;      PRT;   209 AA.
AC   Q9PIA9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Cj0394c.
GN   CJ0394C.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC   Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
DR   EMBL; AL139075; CAB74230.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   209 AA;  23935 MW;  C49313AA9E3D2D19 CRC64;

  Query Match             34.0%;  Score 391.5;  DB 16;  Length 209;
  Best Local Similarity   39.2%;  Pred. No. 6.5e-27;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::||||||:  :|  : : |:   :          :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
                :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203


RESULT 4
Q9WZY5
ID   Q9WZY5      PRELIMINARY;      PRT;   246 AA.
AC   Q9WZY5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TM0883.
GN   TM0883.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
```

```
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001754; AAD35964.1; -.
DR   TIGR; TM0883; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   246 AA;  27154 MW;  9E0309AD462CF266 CRC64;

  Query Match           14.3%;  Score 165;  DB 16;  Length 246;
  Best Local Similarity   23.2%;  Pred. No. 9.7e-07;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:          |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 ----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                      :: |:||   |  :|     |   | || :           :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|       ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

Qy        144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
               |  |   |: | | : :  |:|         :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
               | :  | .|   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 5
Q9HWC1
ID   Q9HWC1      PRELIMINARY;      PRT;   248 AA.
AC   Q9HWC1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA4279.
GN   PA4279.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07667.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001230; Prenyl_site.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   248 AA;  26758 MW;  609D37480899DA3B CRC64;

  Query Match           13.8%;  Score 159.5;  DB 16;  Length 248;
  Best Local Similarity   28.6%;  Pred. No. 3e-06;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
              :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::
Db         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                    :|:| |:|:| ||:  :| |||| | ||:              :|  |  : |
Db        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
               || :|          | :||  | :| :         :|| : |  : :|:| ||||:
Db        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy        191 YLSAFLPHSVCKERLVFDGMEIA 213
               :  |  :     ||| |: :|
Db        222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 6
Q8YQD7
ID   Q8YQD7      PRELIMINARY;      PRT;   276 AA.
AC   Q8YQD7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Alr3896.
GN   ALR3896.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003594; BAB75595.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   276 AA;  30272 MW;  2AD04CA693E56F25 CRC64;

  Query Match           13.7%;  Score 158;  DB 16;  Length 276;
  Best Local Similarity   25.1%;  Pred. No. 4.7e-06;
  Matches   70;  Conservative   51;  Mismatches   94;  Indels   64;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLC-DIGNTRIHFA-----------------QNYQLFSSAKEDLKR 41
              | ::||   ::|  |  :|||:|:|:|                     |   : :  |:
Db          1 MISQQSAKHIENQWLALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEV 60

Qy         42 LGIQKEIF--------------YISVNEENEKALLNCYPNAKNIA-GFFHLETDYVGLG 85
                 :||||              :|:     :  |   | | : |      |   |  ||
Db         61 GSEEKEIFFTFFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPTLG 120

Qy         86 IDRQMACLAVNNG-----VVVDAGSAITIDLIKEGKHL-GGCILPGLA-QYIHAYKKSAK 138
              ||| :|            :|:|||:|:|      ||:| || ||||:  |:    ::: :
Db        121 IDRALALWGAGMSWGFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQ 180

Qy        139 ILEQPFKALDSLEVLPK---STRDAVNYGMVLSVIACIQ-------HLAKNQKIYLCGGD 188
               : :   :|: ||   |:   :| :|:  |:: ::|| ::        |  : |: : |||
Db        181 LPQVEMEAIKSLP--PRFALNTTEAIQSGVIYTLIAGMRDFTEEWLSLFPDGKVAIKGGD 238

Qy        189 ----AKYLSAFLP----HSVCKERLVFDGMEIALKKAGI 219
                  || |  |      : :  |:| ||:  :  ||:
Db        239 RILLLNYLQALYPDLAARLIVEPNLIFWGMQTIV--AGV 275


RESULT 7
P74045
```

```
ID   P74045      PRELIMINARY;      PRT;   257 AA.
AC   P74045;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein slr0812.
GN   SLR0812.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
DR   EMBL; D90911; BAA18120.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   257 AA;  27702 MW;  4291E51EE91FB4A2 CRC64;

  Query Match             12.3%;  Score 141.5;  DB 16;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 0.00012;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy         18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
              |||:|:|:|     : |    |:                ||    :    | |: |
Db         25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy         59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                        |:| ||:                |||| :| |          :|||
Db         80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy        104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
              |:|:||    : | | || |||||   :       ::   |   :| |  ||      |
Db        124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy        159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
               |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||       : |
Db        182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

          206 VFDGM 210
              :| ||
Db        242 IFLGM 246


RESULT 8
Q8RFE4
ID   Q8RFE4      PRELIMINARY;      PRT;   256 AA.
AC   Q8RFE4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bvg accessory factor.
GN   FN0761.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010586; AAL94957.1; -.
KW   Complete proteome.
SQ   SEQUENCE   256 AA;  28601 MW;  CDF681127F08914B CRC64;

  Query Match             12.1%;  Score 139;  DB 16;  Length 256;
  Best Local Similarity   25.9%;  Pred. No. 0.00021;
  Matches   59;  Conservative   31;  Mismatches   86;  Indels   52;  Gaps    8;

Qy         17 DIGNTRI---HFAQNYQLFSSAKEDLKRLGIQKEIF----------YISVNEENEKALLN 63
              ||||| |    :  | :| |: :       : | |          ||: : :   : :
Db          6 DIGNTHIVTGIYDNNGELISTFRIATNDKMTEDEYFSYFNNITKYNEISIKKVDAILISS 65

Qy         64 CYPN--------------------AKNIAGFFHLETDYVGLGIDR------QMACLAV 95
              ||                      | :   |    :| | | ||       |
Db         66 VVPNIIITFQFFARKYFKVEATIVDLEKKLPFTFAKGINYTGFGADRIIDITEAMQKYPD 125

Qy         96 NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQPFKALDSLEVLP 154
              | |: | |:| | |::|:| ::|| ||||:   |:| |  :||:    |    |  ||
Db        126 KNLVIFDFGTATTYDVLKKGVYIGGGILPGIDMSINALYGNTAKLPRVKFTTPSS--VLG 183

Qy        155 KSTRDAVNYGMVLSVIACIQHLAK------NQKIYL--CGGDAKYLSA 194
              |   :   :        |:|: |      |::|::   ||  | |||
Db        184 TDTMKQIQAAIFFGYAGQIKHIIKKINEELNEEIFVLATGGLGKILSA 231


RESULT 9
O67753
ID   O67753      PRELIMINARY;      PRT;   229 AA.
AC   O67753;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein AQ_1924.
GN   AQ_1924.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
DR   EMBL; AE000763; AAC07720.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   229 AA;  26068 MW;  64C347A285FABE9A CRC64;

  Query Match             12.0%;  Score 138.5;  DB 16;  Length 229;
  Best Local Similarity   26.5%;  Pred. No. 0.0002;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy         17 DIGNTRIHFA------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
              |:||: :  |              : :        || ||       ||| :   : :
Db          7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy         64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                |  | :    | :: ||     || ||     :       | ||: ||:|: |||: |
Db         60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy        115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
              ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
```

```
Db        120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172

Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                  : | :   : : | : |: : ||: || | |     :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219


RESULT 10
Q8XHL5
ID   Q8XHL5      PRELIMINARY;      PRT;   259 AA.
AC   Q8XHL5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CPE2468.
GN   CPE2468.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003194; BAB82174.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28819 MW;  0D5FEA3B7A145E10 CRC64;

  Query Match             11.0%;  Score 126.5;  DB 16;  Length 259;
  Best Local Similarity   22.8%;  Pred. No. 0.0027;
  Matches   63;  Conservative   43;  Mismatches   77;  Indels   93;  Gaps   13;

Qy         13 LVLCDIGNTR----IHFAQNY---------------------QLFSSAK---EDLKRLG 43
              ::| |:|||     ||  : |                      |||: ||   ||:: :
Db          2 ILLIDVGNTNIVLGIHDNEKYIASWRISTDSKKTSDEYSIQVMQLFNQAKLNPEDVEGII 61

Qy         44 IQKEIFYISVNEENEKALLNC----------------YPNAKNIAGFFHLETDYVGLGI 86
              |  :  |  : ||      |                | | | :             |
Db         62 ISSVVPNIMHSLENMVRKCFCKEPIVVGPGIKTGINIKYDNPKEV------------GA 108

Qy         87 DRQMACLAV-----NNGVVVDAGSAITIDLIKE-GKHLGGCILPGLAQYIHA-YKKSAKI 139
              || :  :|         :::| |:| |   | | | :||| | ||:     | ::::||:
Db        109 DRIVNAVAAFEKHKKPMIIIDFGTATTFCAITEKGDYLGGNICPGIQISADALFERAAKL 168

Qy        140 ----LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQK------------IY 183
                ||:|        |: |:|  ::  |::   |  :::: |  |              :
Db        169 PRIELEKP------KSVICKNTVTSMQAGIIYGYIGKVEYIVKRMKKEMMDLGEKEPFVL 222

Qy        184 LCGGDAKYLSAFLPHSVCKE---RLVFDGMEIALKK 216
                || | :  :     |  |    :| .:|::|  :|
Db        223 ATGGLAKLV--YSETDVIDEVDRKLTLEGLKILYEK 256


RESULT 11
Q97EB4
ID   Q97EB4      PRELIMINARY;      PRT;   273 AA.
AC   Q97EB4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Predicted transcriptional regulator, homolog of Bvg accessory
DE   factor.
GN   CAC3200.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007815; AAK81136.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30331 MW;  DE7B4D5923E72842 CRC64;

  Query Match             10.2%;  Score 117.5;  DB 16;  Length 273;
  Best Local Similarity   21.4%;  Pred. No. 0.018;
  Matches   57;  Conservative   48;  Mismatches   94;  Indels   67;  Gaps   11;

Qy         10 LKNLVLCDIGNTRI--------HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKAL 61
              || ::: |:||| |             ::| :        |||  :  :   :: :  |
Db         10 LKVILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQ--VMNLFQQDKLDPTL 67

Qy         62 LN-----------CYPNAKNIAGFFHLETDYVGLGI---------------DRQMACLA 94
              :              |     |  :| :    || ||                 || :  :|
Db         68 VEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVA 127

Qy         95 VN-----NGVVVDAGSAITIDLIKE-GKHLGGCILPGLAQYIHA-YKKSAKI----LEQP 143
              :     : :::| |:| |   ::| | :||| | ||:     | ::|:||:    | :|
Db        128 AHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKP 187

Qy        144 FKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQK------------IYLCGGDAKY 191
              |      : |:|  ::  |:|   |  :::: :  |              :   || ||
Db        188 AYA------ICKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKL 241

Qy        192 LSAFLPH-SVCKERLVFDGMEIALKK 216
              :|    :  |    |  :|: |  :|
Db        242 ISEEAKNVDVINPFLTLEGLRIIYEK 267


RESULT 12
Q9JXF1
ID   Q9JXF1      PRELIMINARY;      PRT;   592 AA.
AC   Q9JXF1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein/Bvg accessory factor.
GN   NMB2075.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
```

```
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002557; AAF42394.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; NMB2075; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  64701 MW;  9E27421DA2B41DE3 CRC64;

  Query Match           10.1%;  Score 116.5;  DB 16;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.059;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 13
Q9JWI7
ID   Q9JWI7      PRELIMINARY;      PRT;   592 AA.
AC   Q9JWI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein (EC 6.3.4.15).
GN   BIRA OR NMA0357.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162752; CAB83659.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   592 AA;  64468 MW;  7ED2E0CD3B31C630 CRC64;

  Query Match           10.1%;  Score 116.5;  DB 16;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.059;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 14
Q98Q93
ID   Q98Q93      PRELIMINARY;      PRT;   224 AA.
AC   Q98Q93;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Hypothetical protein MYPU_4730.
GN   MYPU_4730.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445564; CAC13646.1; -.
DR   MypuList; MYPU_4730; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   224 AA;  25363 MW;  240600CED6AB2B4B CRC64;

  Query Match           10.1%;  Score 116;  DB 16;  Length 224;
  Best Local Similarity   22.3%;  Pred. No. 0.019;
  Matches   51;  Conservative   30;  Mismatches   94;  Indels   54;  Gaps    7;

Qy         14 VLCDIGNTRIHFA----------------------QNYQLFSSAKEDLKRLGIQKEI 48
              :  |:||:   ||                        ::: ||   | |:      |||
Db          3 IFIDLGNSFAKFALKNENKAHFLFRLKTQSVVDPSFEIKSFNLFEFNKLDV------KEI 56

Qy         49 FYISVNEENEKALL-----NCYPNAKNIAGFFHLE----------TDYVGLGI--DRQM 90
```

```
                 ||    |  :|     : : ||| |  | |             |  :|| |  :
Db         57 LICSVRNAKENQILEFKLKSIFKNAK-IDFFIHKSQSLVKFCQKELTSEIGLDIVANAYY 115

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
                  | : :  |:|| |  ||     |  | ||: |    :       :|  |  :
Db        116 VLHKSKNAIFISLGTAIVITQIKNSSIEGISIYPGIYQSFKNFFNVVAKIESNFNIIKIP 175

Qy        151 EVLPKSTRDAVN----YGMVLSVIACIQHLAKNQKIYLCGGDAKYLSAF 195
              :| |:| ::::     | |  :   |  : |   |:: ||   :   |
Db        176 PILGKTTLESISSSLVRGSVFLLKGYIDEIDKTSDIFITGGTITFSKIF 224


RESULT 15
Q9PCI4
ID   Q9PCI4      PRELIMINARY;      PRT;   242 AA.
AC   Q9PCI4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Xf1795.
GN   XF1795.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004001; AAF84603.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   242 AA;  25658 MW;  EDC5DE7C0449C1E9 CRC64;

  Query Match             10.0%;  Score 115.5;  DB 16;  Length 242;
  Best Local Similarity   31.5%;  Pred. No. 0.023;
  Matches   45;  Conservative   22;  Mismatches   55;  Indels   21;  Gaps    6;

Qy         85 GIDRQMACL---AVNNGVVVDAGSAITIDLI-KEGKHLGGCILPG---LAQYIHAYKKSA 137
              |:|| :| |      | :||  |:|:|||:   | |||| |      : | :|     |
Db         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152

Qy        138 KILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACI--------QHLAKNQKIYLCGGDA 189
             :  : |    : ||   : | ||:  |   : :| |        | | :: :: | ||

Db        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211

Qy        190 KYLSAFLPHSVCKERLVFDGMEI 212
                 |  :|   | :  || ||: |
Db        212 ASLLPWLGDVVHRPTLVLDGLAI 234


Search completed: June 24, 2003, 21:59:32
Job time : 20.1879 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 3.87859 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-14
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	109.5	9.5	233	1	YACB_BACSU	P37564 bacillus su
2	89.5	7.8	232	1	PYRK_ARCFU	O29722 archaeoglob
3	89.5	7.8	1581	1	ACC8_CRICR	Q09427 cricetus cr
4	88.5	7.7	1182	1	DP3A_RICPR	O05974 rickettsia
5	88.5	7.7	1581	1	ACC8_RAT	Q09429 rattus norv
6	87.5	7.6	344	1	HOLA_HAEIN	P43747 haemophilus
7	85.5	7.4	426	1	GLPC_HAEIN	P43801 haemophilus
8	84	7.3	764	1	SYFB_HELPJ	Q9zkf8 helicobacte
9	83	7.2	303	1	HIS1_HAEIN	P43853 haemophilus
10	82.5	7.2	349	1	GPDA_DROAE	Q27556 drosophila
11	82.5	7.2	352	1	GPDA_DROVI	P07735 drosophila
12	82.5	7.2	424	1	PSGA_HUMAN	Q15235 homo sapien
13	81.5	7.1	293	1	NPL_HAEIN	P44539 haemophilus
14	81.5	7.1	314	1	ARCC_CLOPE	Q46171 clostridium
15	81	7.0	1580	1	ACC8_HUMAN	Q09428 homo sapien
16	81	7.0	2452	1	RPB1_PLAFD	P14248 plasmodium
17	80.5	7.0	353	1	GPDA_DROPS	Q27928 drosophila
18	80	6.9	253	1	YTBQ_BACSU	P53560 bacillus su
19	80	6.9	327	1	MTH1_HAEHA	P05102 haemophilus
20	80	6.9	449	1	MURF_RICCN	Q92h60 rickettsia
21	80	6.9	558	1	VB04_VACCC	P21001 vaccinia vi
22	80	6.9	558	1	VB04_VACCV	P24769 vaccinia vi
23	80	6.9	633	1	YR45_CAEEL	Q09562 caenorhabdi
24	79.5	6.9	370	1	ALR2_SCHPO	Q9p5n3 schizosacch
25	79	6.9	310	1	ARCL_ECOLI	Q46807 escherichia
26	79	6.9	374	1	ADHA_HUMAN	P07327 homo sapien
27	79	6.9	773	1	GYRB_HELPJ	Q9zlx3 helicobacte
28	79	6.9	1065	1	CARB_CLOAB	Q97ft3 clostridium
29	79	6.9	2799	1	EDD_HUMAN	O95071 homo sapien
30	78.5	6.8	131	1	YIL2_STAAU	P41370 staphylococ
31	78.5	6.8	349	1	GPDA_DROEZ	Q27567 drosophila
32	78.5	6.8	355	1	ALR_RICCN	Q92jd9 rickettsia
33	78.5	6.8	359	1	GPDA_DROKA	O97463 drosophila
34	78.5	6.8	500	1	CPN1_CAVPO	Q64408 cavia porce
35	78	6.8	425	1	CLPX_RICCN	Q92gq4 rickettsia
36	78	6.8	717	1	NTP2_MSEPV	Q9yw06 melanoplus
37	78	6.8	920	1	EDD_RAT	Q62671 rattus norv
38	77.5	6.7	1448	1	UGGG_SCHPO	Q09140 schizosacch
39	77	6.7	389	1	PSD6_ORYSA	Q8w425 oryza sativ
40	77	6.7	917	1	VGLB_HSVB2	P12641 bovine herp
41	76.5	6.6	326	1	Y338_CHLMU	Q9pkx2 chlamydia m
42	76.5	6.6	335	1	G3P_BORBU	P46795 borrelia bu
43	76.5	6.6	484	1	MURE_CLOPE	Q8xj99 clostridium
44	76.5	6.6	508	1	YI57_THEMA	Q9x2h2 thermotoga
45	76.5	6.6	538	1	THSB_METTH	O26885 methanobact

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             9.5%;  Score 109.5;  DB 1;  Length 233;
  Best Local Similarity   26.7%;  Pred. No. 0.021;
  Matches   35;  Conservative   24;  Mismatches   47;  Indels   25;  Gaps    5;

Qy         74 FFHLETDYVG----------------LGIDRQMACLAV-----NNGVVVDAGSAITIDLI 112
              :||:|   ||                :| || :  :|      |  :||| |:| |   |
Db         80 YFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYI 139

Qy        113 KEGK-HLGGCILPGLAQYIHA-YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVI 170
               | | ::|| | ||:     | | ::||:        |:  :: |:|  |:  |::   :
Db        140 DENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDN--IIGKNTVSAMQSGILFGYV 197

Qy        171 ACIQHLAKNQK 181
               :: : |  |
Db        198 GQVEGIVKRMK 208


RESULT 2
PYRK_ARCFU
ID   PYRK_ARCFU     STANDARD;      PRT;   232 AA.
AC   O29722;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable dihydroorotate dehydrogenase electron transfer subunit.
GN   PYRK OR AF0528.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- FUNCTION: Is responsible for channelling the electrons from the
CC       oxidation of dihydroorotate from the FMN redox center in the pyrD
CC       subunit to the ultimate electron acceptor NAD(+) (By similarity).
CC   -!- COFACTOR: Binds a 2Fe-2S cluster and FAD (By similarity).
CC   -!- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC   -!- SUBUNIT: Heterotetramer of 2 pyrK and 2 pyrD subunits (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE PYRK FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001068; AAB90709.1; -.
DR   TIGR; AF0528; -.
DR   InterPro; IPR000564; 2Fe2S_ferredoxin.
DR   InterPro; IPR001834; Cyt_B5_reductase.
DR   InterPro; IPR001433; Oxred_FAD/NAD(P).
DR   Pfam; PF00175; NAD_binding; 1.
DR   Pfam; PF00970; FAD_binding_6; 1.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW   Pyrimidine biosynthesis; Transport; Electron transport; Metal-binding;
KW   Iron; Iron-sulfur; Flavoprotein; FAD; Complete proteome.
FT   METAL       202    202       IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
FT   METAL       207    207       IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
FT   METAL       210    210       IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
FT   METAL       219    219       IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
SQ   SEQUENCE   232 AA;  25639 MW;  907C6A5D3098784B CRC64;

  Query Match             7.8%;  Score 89.5;  DB 1;  Length 232;
  Best Local Similarity   24.7%;  Pred. No. 1.1;
  Matches   54;  Conservative   36;  Mismatches   74;  Indels   55;  Gaps   14;

Qy         39 LKRLGIQKEIFYISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNG 98
              ::||  :    | |::       |  ||      | | :   ||  | :     |: :|
Db          9 IERLTDEVATLYFSIS-------LRSYP------GQFVMV--YVP-GCEEIPLSLSSSNS 52

Qy         99 VVVDAGSAITIDLI--KEGKHL------GGCILP----------GLA--QYIHAY---- 133
              | | |    |  ||  |||:::      |    |           |:|  :|:: |
Db         53 VTVKAVGETTASLINAKEGQYVGVRGAFGSAFTPSKRALIVAGGIGIAPMKYLYEYLMKC 112

Qy        134 --KKSAKILEQPFKA---LDSLEVLPKSTRDAVNYGMVLSVIAC--IQHLAKNQKIYLCG 186
                | |    |: |    ||  : :  :| |  ::|:  :|:    :: | : :|||:||
Db        113 GSKVSVVYGERTAKNLFWLDKFDRITVTTEDG-SFGLKGTVLDALKLEKLDEYEKIYVCG 171

Qy        187 GDAKYLSAFLPHSVCKERLVFDGMEIALK---KAGILEC 222
               :    : :      |||   |  | :|:   : ||  |
Db        172 SEGMLRATY---DFLKERDALDKAEFSLERYMRCGIGVC 207


RESULT 3
ACC8_CRICR
ID   ACC8_CRICR     STANDARD;      PRT;  1581 AA.
AC   Q09427;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sulfonylurea receptor 1.
GN   ABCC8 OR SUR.
OS   Cricetus cricetus (Black-bellied hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetus.
OX   NCBI_TaxID=10034;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=95232532; PubMed=7716547;
RA   Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV,
RA   Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J.,
RA   Nelson D.A.;
RT   "Cloning of the beta cell high-affinity sulfonylurea receptor: a
RT   regulator of insulin secretion.";
RL   Science 268:423-426(1995).
RN   [2]
RP   CARBOHYDRATE-LINKAGE SITE ASN-9.
```

```
RX   MEDLINE=97098100; PubMed=8942641;
RA   Nelson D.A., Bryan J., Wechsler S., Clement J.P. IV, Aguilar-Bryan L.;
RT   "The high-affinity sulfonylurea receptor: distribution, glycosylation,
RT   purification, and immunoprecipitation of two forms from endocrine and
RT   neuroendocrine cell lines.";
RL   Biochemistry 35:14793-14799(1996).
RN   [3]
RP   TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITE ASN-9.
RX   MEDLINE=99436117; PubMed=10506167;
RA   Raab-Graham K.F., Cirilo L.J., Boettcher A.A., Radeke C.M.,
RA   Vandenberg C.A.;
RT   "Membrane topology of the amino-terminal region of the sulfonylurea
RT   receptor.";
RL   J. Biol. Chem. 274:29122-29129(1999).
CC   -!- FUNCTION: PUTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC       POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC       AND INSULIN RELEASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40623; AAA99201.1; -.
DR   HSSP; P13569; 1NBD.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABCtranprtrTM.
DR   InterPro; IPR000388; Surea_receptor.
DR   Pfam; PF00005; ABC_tran; 2.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   PRINTS; PR01092; SULFNYLUREAR.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   Receptor; Transmembrane; ATP-binding; Transport; Glycoprotein; Repeat;
KW   Phosphorylation.
FT   INIT_MET      0      0
FT   DOMAIN        1     33       EXTRACELLULAR.
FT   TRANSMEM     34     54       1.
FT   DOMAIN       55     74       CYTOPLASMIC.
FT   TRANSMEM     75     95       2.
FT   DOMAIN       96    100       EXTRACELLULAR.
FT   TRANSMEM    101    121       3.
FT   DOMAIN      122    133       CYTOPLASMIC.
FT   TRANSMEM    134    153       4.
FT   DOMAIN      154    166       EXTRACELLULAR.
FT   TRANSMEM    167    193       5.
FT   DOMAIN      194    310       CYTOPLASMIC.
FT   TRANSMEM    311    330       6 (BY SIMILARITY).
FT   DOMAIN      331    354       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    355    375       7 (BY SIMILARITY).
FT   DOMAIN      376    433       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    434    454       8 (BY SIMILARITY).
FT   DOMAIN      455    457       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    458    478       9 (BY SIMILARITY).
FT   DOMAIN      479    540       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    541    561       10 (BY SIMILARITY).
FT   DOMAIN      562    583       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    584    604       11 (BY SIMILARITY).
FT   DOMAIN      605   1004       CYTOPLASMIC.
FT   TRANSMEM   1005   1025       12.
FT   DOMAIN     1026   1072       EXTRACELLULAR.
FT   TRANSMEM   1073   1093       13 (BY SIMILARITY).
FT   DOMAIN     1094   1137       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1138   1158       14 (BY SIMILARITY).
FT   DOMAIN     1159   1159       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1160   1180       15 (BY SIMILARITY).
FT   DOMAIN     1181   1251       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1252   1272       16 (BY SIMILARITY).
FT   DOMAIN     1273   1276       EXTRACELLULAR.
FT   TRANSMEM   1277   1297       17.
FT   DOMAIN     1298   1581       CYTOPLASMIC.
FT   NP_BIND     712    719       ATP (POTENTIAL).
FT   NP_BIND    1378   1385       ATP (POTENTIAL).
FT   CARBOHYD      9      9       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .).
SQ   SEQUENCE   1581 AA;  177015 MW;  0EFCCFB3A454E0B1 CRC64;

  Query Match             7.8%;  Score 89.5;  DB 1;  Length 1581;
  Best Local Similarity   22.7%;  Pred. No. 9.6;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1280

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1281 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1336

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
               :  |         :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1337 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1394

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1395 VDMFEGRIIIDGIDIA 1410


RESULT 4
DP3A_RICPR
ID   DP3A_RICPR     STANDARD;      PRT;  1182 AA.
AC   O05974;
DT   15-DEC-1998 (Rel. 37, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA polymerase III alpha subunit (EC 2.7.7.7).
GN   DNAE OR RP778.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE OF 1-780 FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=97419517; PubMed=9274032;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Genomic rearrangements during evolution of the obligate
RT   intracellular parasite Rickettsia prowazekii as inferred from an
RT   analysis of 52015 bp nucleotide sequence.";
RL   Microbiology 143:2783-2795(1997).
RN   [3]
RP   SEQUENCE OF 822-1182 FROM N.A.
RC   STRAIN=Madrid E, and B;
RX   MEDLINE=99416441; PubMed=10486973;
RA   Andersson J.O., Andersson S.G.E.;
```

```
RT   "Genome degradation is an ongoing process in Rickettsia.";
RL   Mol. Biol. Evol. 16:1178-1191(1999).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC       THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC       CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC       DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC       THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC       AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235273; CAA15204.1; -.
DR   EMBL; Y11785; CAA72479.1; -.
DR   EMBL; AJ238755; CAB56087.1; -.
DR   EMBL; AJ238756; CAB56091.1; -.
DR   InterPro; IPR004013; PHP_C.
DR   InterPro; IPR003141; PHP_N.
DR   InterPro; IPR004805; PolC_alpha.
DR   InterPro; IPR004365; tRNA_anti.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   Pfam; PF02231; PHP_N; 1.
DR   Pfam; PF02811; PHP_C; 1.
DR   SMART; SM00481; POLIIIAc; 1.
DR   TIGRFAMs; TIGR00594; polc; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   Complete proteome.
FT   CONFLICT    376    376       I -> V (IN REF. 2).
SQ   SEQUENCE   1182 AA;  133462 MW;  0FE9B56969C2E52A CRC64;

  Query Match             7.7%;  Score 88.5;  DB 1;  Length 1182;
  Best Local Similarity   21.3%;  Pred. No. 8.4;
  Matches   59;  Conservative   35;  Mismatches  104;  Indels   79;  Gaps    9;

Qy          4 RQSFTDLKNLVLCDIG-NTRIHFA----------------QNYQLFSSAKED-LKRLGI 44
              |: |:|| | :   |    | :||                 : | :   | |  | ||
Db        243 RKLFSDLPNAIQNTINLRERCYFAAHPNPPMLPNFSTQDISETYLIRKYATEGLLARLVT 302

Qy         45 QKEIFYISVNEENEKALLNCYPNAK---------NIAGFFHLETDYVGLGIDRQMACLAV 95
              : :   ||: |  ||      :             | ||:| : :|::     :   : |
Db        303 KFKAENISL-EHQEKLKTEYFTRLNYELDIICNMNFAGYFLIVSDFIKWS---KKQGILV 358

Qy         96 NNGVVVDAGSAIT-------IDLIKEG-------------------------------- 115
                |    ||| :         :| || |
Db        359 GPGRGSGAGSVVAWSLLITDLDPIKFGLLFERFLNPERISMPDFDIDFCQERREEVINYV 418

Qy        116 ------KHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL-EVLPKSTRDAVNYGMVLS 168
                   :|  |  |  |      |  |::|  |:|  | | |::| |  : |:    :
Db        419 RSKYGPNRVGQIITFGKMQAKAVIKDVARVLSLPYKLADYLTELVPFSAINPVSLEQAIR 478

Qy        169 VIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERL 205
               :  : : ||   :|   |||: :   |  |:  | |
Db        479 EVPELANAAKGNGLYNLEGDAELIKLVLDTSLILEGL 515


RESULT 5
ACC8_RAT
ID   ACC8_RAT       STANDARD;      PRT;  1581 AA.
AC   Q09429; O54989; P70532; Q9EQT0;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sulfonylurea receptor 1.
GN   ABCC8 OR SUR1 OR SUR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM C).
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=95232532; PubMed=7716547;
RA   Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV,
RA   Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J.,
RA   Nelson D.A.;
RT   "Cloning of the beta cell high-affinity sulfonylurea receptor: a
RT   regulator of insulin secretion.";
RL   Science 268:423-426(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RC   TISSUE=Pancreatic islets;
RA   Faure C., Partiseti M., Gouhier C., Graham D.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM B).
RC   TISSUE=Pancreatic islets;
RA   Blache P., Peyrollier K., Gros L., Bataille D.;
RT   "Identification of two forms of sulfonylurea receptors (SUR 1A and SUR
RT   1B) by means of polymerase chain reaction, subcloning, DNA sequencing
RT   and tissue distribution.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RC   TISSUE=Vascular smooth muscle;
RA   Cao K., Wang R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   CARBOHYDRATE-LINKAGE SITE ASN-9.
RX   MEDLINE=97098100; PubMed=8942641;
RA   Nelson D.A., Bryan J., Wechsler S., Clement J.P. IV, Aguilar-Bryan L.;
RT   "The high-affinity sulfonylurea receptor: distribution, glycosylation,
RT   purification, and immunoprecipitation of two forms from endocrine and
RT   neuroendocrine cell lines.";
RL   Biochemistry 35:14793-14799(1996).
CC   -!- FUNCTION: PUTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC       POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC       AND INSULIN RELEASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; A (shown here), B and C; may be
CC       produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40624; AAA99237.1; -.
DR   EMBL; X97279; CAA65934.1; -.
DR   EMBL; AF039595; AAB96684.1; -.
DR   EMBL; AB052294; BAB19011.1; -.
DR   HSSP; P13569; 1NBD.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABCtranprtrTM.
DR   InterPro; IPR000388; Surea_receptor.
DR   Pfam; PF00005; ABC_tran; 2.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   PRINTS; PR01092; SULFNYLUREAR.
DR   ProDom; PD000006; ABC_transportr; 2.
```

```
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   Receptor; Transmembrane; ATP-binding; Transport; Glycoprotein; Repeat;
KW   Phosphorylation; Alternative splicing.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     33       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM     34     54       1 (BY SIMILARITY).
FT   DOMAIN       55     74       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     75     95       2 (BY SIMILARITY).
FT   DOMAIN       96    100       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    101    121       3 (BY SIMILARITY).
FT   DOMAIN      122    133       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    134    153       4 (BY SIMILARITY).
FT   DOMAIN      154    167       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    168    193       5 (BY SIMILARITY).
FT   DOMAIN      194    310       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    311    330       6 (BY SIMILARITY).
FT   DOMAIN      331    354       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    355    375       7 (BY SIMILARITY).
FT   DOMAIN      376    433       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    434    454       8 (BY SIMILARITY).
FT   DOMAIN      455    457       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    458    478       9 (BY SIMILARITY).
FT   DOMAIN      479    540       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    541    561       10 (BY SIMILARITY).
FT   DOMAIN      562    583       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    584    604       11 (BY SIMILARITY).
FT   DOMAIN      605   1004       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1005   1025       12 (BY SIMILARITY).
FT   DOMAIN     1026   1072       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1073   1093       13 (BY SIMILARITY).
FT   DOMAIN     1094   1137       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1138   1158       14 (BY SIMILARITY).
FT   DOMAIN     1159   1159       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1160   1180       15 (BY SIMILARITY).
FT   DOMAIN     1181   1251       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1252   1272       16 (BY SIMILARITY).
FT   DOMAIN     1273   1276       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1277   1297       17 (BY SIMILARITY).
FT   DOMAIN     1298   1581       CYTOPLASMIC (BY SIMILARITY).
FT   NP_BIND     712    719       ATP (POTENTIAL).
FT   NP_BIND    1378   1385       ATP (POTENTIAL).
FT   CARBOHYD      9      9       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT   VARSPLIC    740    740       MISSING (IN ISOFORM C).
FT   VARSPLIC   1251   1288       MISSING (IN ISOFORM B).
FT   CONFLICT    486    486       S -> T (IN REF. 1).
FT   CONFLICT    698    698       I -> T (IN REF. 4).
FT   CONFLICT    835    836       QR -> PG (IN REF. 1).
FT   CONFLICT   1313   1313       R -> G (IN REF. 1).
SQ   SEQUENCE   1581 AA;  177053 MW;  C554B1A48B2D3901 CRC64;

  Query Match             7.7%;  Score 88.5;  DB 1;  Length 1581;
  Best Local Similarity   21.2%;  Pred. No. 12;
  Matches   56;  Conservative   48;  Mismatches   71;  Indels   89;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNNGVVVDAGSAITID 110
                  | :: |||  |         :. :|:|       ||:     |:: | ::|:
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIG-------ACV-----VLIAAATSISNS 1270

Qy        111 LIKEGKHLGGCILPGL------AQYIH-------------AYKKSAKILE---QPFKAL 147
              | :|     | :  ||      : |::             | |:   :|:   : :: |
Db       1271 LHRELS--AGLVGLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHTLLKTEAESYEGL 1328

Qy        148 DSLEVLPKSTRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK-- 190
               :  ::||:  |         :| |   |  :  ::|    ::  ||| :|| |  |
Db       1329 LAPSLIPKNWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSS 1386

Qy        191 -YLSAFLPHSVCKERLVFDGMEIA 213
```

```
                |: |    : : |:: ||::||
Db       1387 FSLAFFRMVDMFEGRIIIDGIDIA 1410


RESULT 6
HOLA_HAEIN
ID   HOLA_HAEIN     STANDARD;      PRT;   344 AA.
AC   P43747;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA polymerase III, delta subunit (EC 2.7.7.7).
GN   HOLA OR HI0923.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC       THE DELTA SUBUNIT SEEMS TO INTERACT WITH THE GAMMA SUBUNIT TO
CC       TRANSFER THE BETA SUBUNIT ON THE DNA (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC       CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC       DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC       THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC       AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32774; AAC22583.1; -.
DR   TIGR; HI0923; -.
DR   TIGRFAMs; TIGR01128; holA; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   Complete proteome.
SQ   SEQUENCE   344 AA;  39845 MW;  5048DF5A42A97B5F CRC64;

  Query Match             7.6%;  Score 87.5;  DB 1;  Length 344;
  Best Local Similarity   24.2%;  Pred. No. 2.5;
  Matches   29;  Conservative   24;  Mismatches   44;  Indels   23;  Gaps    5;

Qy         43 GIQKEIFYISVNE-ENEKALLNCY-PNAKNIAGFFHLETDYVGLGIDR---QMACLAVNN 97
              ||:|: ::|::|: |    |:||  |  :|:  :    |  :||  |    |  | : |
Db       119 GIEKQTWFITLNQYEPNTILINCQTPTVENLPRWVKNRTKAMGLDADNEAIQQLCYSYEN 178

Qy         98 GVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQ-----PFKALDSLEV 152
                 ::    :   :||:     |               | :   ::||     ||: :|:| |
Db       179 NLLALKQALQLLDLLYPDHKLN------------YNRVISVVEQSSIFTPFQWIDALLV 225
```

```
RESULT 7
GLPC_HAEIN
ID   GLPC_HAEIN     STANDARD;      PRT;   426 AA.
AC   P43801;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Anaerobic glycerol-3-phosphate dehydrogenase subunit C (G-3-P
DE   dehydrogenase).
GN   GLPC OR HI0683.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: ELECTRON TRANSFER PROTEIN; MAY ALSO FUNCTION AS THE
CC       MEMBRANE ANCHOR FOR THE GLPAB DIMER (BY SIMILARITY).
CC   -!- PATHWAY: FIRST STEP IN THE TRANSFER OF REDUCING EQUIVALENTS
CC       FROM G-3-P TO A SHORT ELECTRON TRANSFER CHAIN TERMINATING WITH
CC       FUMARATE AS THE ULTIMATE ELECTRON ACCEPTOR.
CC   -!- SUBUNIT: COMPOSED OF A CATALYTIC GLPA/B DIMER AND OF GLPC
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE CYTOPLASMIC MEMBRANE
CC       OFTEN OCCURRING IN VESICLES ASSOCIATED WITH FUMARATE REDUCTASE
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC       'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32751; AAC22343.1; -.
DR   TIGR; HI0683; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR004017; DUF224.
DR   Pfam; PF00037; fer4; 1.
DR   Pfam; PF02754; DUF224; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Membrane; Complete proteome.
FT   METAL        32     32       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        35     35       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        38     38       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        42     42       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        79     79       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        82     82       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        85     85       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        89     89       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   426 AA;  48068 MW;  A6EA5F33C6C4FC3C CRC64;

  Query Match             7.4%;  Score 85.5;  DB 1;  Length 426;
  Best Local Similarity   30.4%;  Pred. No. 4.8;
  Matches   34;  Conservative   17;  Mismatches   40;  Indels   21;  Gaps    6;

Qy          56 ENEKA-----LLNCYPN-AKNIAGFFHLETDYVGLGIDRQMACLAVNNGV-VVDAGSAIT 108
               | ||      ::| :|| |:||| |    |||:|  :|         ||: |:   |: :
Db         226 EKEKCCGLPLMVNGFPNRARNIAQF---NTDYIGKMVD--------ENGIDVISEASSCS 274

Qy         109 IDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDA 160
               ::|  |  |: |     :  :||       | | ||   :| : |   | |
Db         275 LNLRDEYHHILGIDNAKVRPHIHMV---TPFLYQLFKEGKTLPLKPLKLRVA 323


RESULT 8
SYFB_HELPJ
ID   SYFB_HELPJ     STANDARD;      PRT;   764 AA.
AC   Q9ZKF8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE   tRNA ligase beta chain) (PheRS).
GN   PHET OR JHP0979.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC       diphosphate + L-phenylalanyl-tRNA(Phe).
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC       CHAIN FAMILY. SUBFAMILY 1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use, by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001527; AAD06558.1; -.
DR   InterPro; IPR005146; B3_4.
DR   InterPro; IPR005147; B5.
DR   InterPro; IPR005121; Fdx-AntiCB.
DR   InterPro; IPR004532; PheT_bact.
DR   InterPro; IPR002547; tRNA_bind.
DR   Pfam; PF01588; tRNA_bind; 1.
DR   Pfam; PF03147; FDX-ACB; 1.
DR   Pfam; PF03483; B3_4; 1.
DR   Pfam; PF03484; B5; 1.
DR   TIGRFAMs; TIGR00472; pheT_bact; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
SQ   SEQUENCE   764 AA;  85154 MW;  70BDAFD202FED425 CRC64;

  Query Match             7.3%;  Score 84;  DB 1;  Length 764;
  Best Local Similarity   21.5%;  Pred. No. 12;
  Matches   53;  Conservative   36;  Mismatches   85;  Indels   72;  Gaps   12;

Qy           2 PARQSFTDLKNLVLC----------DIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYI 51
               |       |:  ::|          ::|   |   :   :::| :|::::||     | :
Db         519 PITTELNTLRTSLVCGLLDASLRNKNLGFKSIALYEKGSVYNSKREEIQKLG-----FLV 573
```

```
Qy          52 SVNEENEKALLNCYPNAK--------------NIAGFFHLE--TDYVGLGIDRQMACLAV 95
               |  :: |      ||:||                | | | ||  |    :    | | :
Db         574 SGLQKKES-----YPHAKGKAWDFYSFAECVSRIIGDFSLEKLTTQTPINHPYQSAKIIQ 628

Qy          96 NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKK-----SAKILEQPFKALDSL 150
               || :         | :| :       | | : | :  ::       |  |::|   |
Db         629 NNEI---------IGVIAK-------IHPKVIQELDLFESYYAEIDASKLKRPAMLLKPF 672

Qy         151 EVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKY--LSAFLPHSVCKERLVFD 208
                : | | ||     :::         : |  |       ||:   ||  ||  : ||
Db         673 SIYPSSVRDLT---LIIDENTAFSKIKKALK------DAQIPNLSEILPLDIFKE----S 719

Qy         209 GMEIAL 214
               |  |||
Db         720 GNTIAL 725


RESULT 9
HIS1_HAEIN
ID   HIS1_HAEIN     STANDARD;      PRT;   303 AA.
AC   P43853;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP phosphoribosyltransferase (EC 2.4.2.17).
GN   HISG OR HI0468.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: 1-(5-phospho-D-ribosyl)-ATP + diphosphate =
CC       ATP + 5-phospho-alpha-D-ribose 1-diphosphate.
CC   -!- PATHWAY: Histidine biosynthesis; first step. Very important in the
CC       regulation of histidine metabolism.
CC   -!- SUBUNIT: Homohexamer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ATP PHOSPHORIBOSYLTRANSFERASE FAMILY.
CC       LONG SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32729; AAC22127.1; -.
DR   TIGR; HI0468; -.
DR   InterPro; IPR001348; HisG.
DR   Pfam; PF01634; HisG; 1.
DR   ProDom; PD003516; HisG; 1.
DR   TIGRFAMs; TIGR00070; hisG; 1.
DR   PROSITE; PS01316; ATP_P_PHORIBOSYLTR; 1.
KW   Histidine biosynthesis; Transferase; Glycosyltransferase;
KW   Complete proteome.

SQ   SEQUENCE   303 AA;  33821 MW;  08C14D1F6E98A31D CRC64;

  Query Match             7.2%;  Score 83;  DB 1;  Length 303;
  Best Local Similarity   23.7%;  Pred. No. 5.3;
  Matches   55;  Conservative   34;  Mismatches   79;  Indels   64;  Gaps   12;

Qy          11 KNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKN 70
               | |   | |: |:                  | : ::  | :| :   : :   ||
Db          98 KKLRQLDFGDCRL----------------SLAVDRDCHYENVKDLANRRIATSYP---- 136

Qy          71 IAGFFHLETDYVGLGIDRQMACLAVNNGVVV--DAGSAITI-DLIKEGKHLGGCILPGLA 127
                    ||   |:        :|| :|  | |   || |  | ||:  |  |      ||
Db         137 -----HLLKRYMNENGVSFKSCL-LNGSVEVAPSAGIAYAICDLVSSGATLEA---NGLK 187

Qy         128 QYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGG 187
               :    |:  | ::::     : ||    ||: |:    :|: |  :|  |::: | |
Db         188 EVDVIYRSKACLIQR----AEPLE----STKQAL-VDKLLTRIQGVQQAAESKYIMLHAP 238

Qy         188 DAKY--LSAFLP-----------------HSVCKERLVFDGMEIALKKAG 218
                 |   ::| ||                 | | :| | :: ||  ||:||
Db         239 KEKLEKITALLPGVENPTILPLASDTTRVAMHVVSQENLFWETME-QLKEAG 289


RESULT 10
GPDA_DROAE
ID   GPDA_DROAE     STANDARD;      PRT;   349 AA.
AC   Q27556;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic (EC 1.1.1.8)
DE   (GPD-C) (GPDH-C).
GN   GPDH.
OS   Drosophila americana (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=40366;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96025618; PubMed=8522168;
RA   Tominaga H., Narise S.;
RT   "Sequence evolution of the Gpdh gene in the Drosophila virilis
RT   species group.";
RL   Genetica 96:293-302(1995).
CC   -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC       phosphate + NADH.
CC   -!- PATHWAY: THIS ENZYME IS ONE OF THE TWO ENZYMES CONSTITUTING THE
CC       ALPHA-GLYCEROPHOSPHATE CYCLE, IN WHICH THIS ENZYME OPERATES IN
CC       CONJUNCTION WITH A MITOCHONDRIAL ALPHA-GLYCEROPHOSPHATE OXIDASE.
CC       THE ALPHA-GLYCEROPHOSPHATE CYCLE IS ESSENTIAL FOR THE PRODUCTION
CC       OF ENERGY FOR FLIGHT IN INSECTS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC       DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D50089; BAA20286.1; -.
DR   FlyBase; FBgn0015636; Dame\Gpdh.
DR   InterPro; IPR001652; NAD_Gly3P_dh.
DR   Pfam; PF01210; NAD_Gly3P_dh; 1.
DR   PRINTS; PR00077; GPDHDRGNASE.
DR   ProDom; PD001649; NAD_Gly3P_dh; 1.
DR   PROSITE; PS00957; NAD_G3PDH; 1.
KW   Oxidoreductase; NAD.
```

```
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   349 AA;  38191 MW;  8D8C2F5C56E0BDB6 CRC64;

  Query Match             7.2%;  Score 82.5;  DB 1;  Length 349;
  Best Local Similarity   25.7%;  Pred. No. 6.8;
  Matches   43;  Conservative   27;  Mismatches   54;  Indels   43;  Gaps   10;

Qy         10 LKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAK 69
              |||:| |  |      |    :|  : |  : |||: : | :: |           || :|
Db        204 LKNIVACGAG-----FVDGLKLGDNTKAAVIRLGLMEMIRFVDV----------FYPGSK 248

Qy         70 NIAGFFH------LETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKE---GKHLGG 120
               :: ||       | |   | | :|:::   | :|  ::       || ||   |: | |
Db        249 -LSTFFESCGVADLITTCYG-GRNRRVSEAFVTSGKTIE-------DLEKEMLNGQKLQG 299

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSL------EVLPKSTRDAV 161
                 |  |: :: |    | ||  |    ::      :: ||   | :
Db        300 ---PPTAEEVN-YMLKNKGLEDKFPLFTAIHKICTNQLKPKDLIDCI 342


RESULT 11
PDA_DROVI
ID   GPDA_DROVI     STANDARD;      PRT;   352 AA.
AC   P07735; Q27634;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic (EC 1.1.1.8)
DE   (GPD-C) (GPDH-C).
GN   GPDH.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7244;
RN   [1]
RP   SEQUENCE.
RA   Arai K., Tominaga H., Yokote Y., Narise S.;
RT   "The complete amino-acid sequence of cytoplasmic glycerol-3-phosphate
RT   dehydrogenase from Drosophila virilis.";
RL   Biochim. Biophys. Acta 953:6-13(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Narise S., Tominaga H.;
RT   "Temperature-dependency differencies in GPDH allozmyes associated
RT   with a single base change in the coding region of the GPDH
RT   structural gene in Drosophila virilis.";
RL   Life Sci. Adv. (Genet.) 11:39-45(1992).
RN   [3]
     SEQUENCE FROM N.A.
RX   MEDLINE=92305069; PubMed=1610907;
RA   Tominaga H., Shiba T., Narise S.;
RT   "Structure of Drosophila virilis glycerol-3-phosphate dehydrogenase
RT   gene and a comparison with the Drosophila melanogaster gene.";
RL   Biochim. Biophys. Acta 1131:233-238(1992).
CC   -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC       phosphate + NADH.
CC   -!- PATHWAY: THIS ENZYME IS ONE OF THE TWO ENZYMES CONSTITUTING THE
CC       ALPHA-GLYCEROPHOSPHATE CYCLE, IN WHICH THIS ENZYME OPERATES IN
CC       CONJUNCTION WITH A MITOCHONDRIAL ALPHA-GLYCEROPHOSPHATE OXIDASE.
CC       THE ALPHA-GLYCEROPHOSPHATE CYCLE IS ESSENTIAL FOR THE PRODUCTION
CC       OF ENERGY FOR FLIGHT IN INSECTS.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC       DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59076; CAA41800.1; -.
DR   EMBL; D10697; BAA01539.1; -.
DR   PIR; JS0023; JS0023.
DR   FlyBase; FBgn0013081; Dvir\Gpdh.
DR   InterPro; IPR001652; NAD_Gly3P_dh.
DR   Pfam; PF01210; NAD_Gly3P_dh; 1.
DR   PRINTS; PR00077; GPDHDRGNASE.
DR   ProDom; PD001649; NAD_Gly3P_dh; 1.
DR   PROSITE; PS00957; NAD_G3PDH; 1.
KW   Oxidoreductase; NAD.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       BLOCKED.
FT   CONFLICT     14     14       W -> N (IN REF. 1).
FT   CONFLICT     34     34       E -> K (IN REF. 2).
SQ   SEQUENCE   352 AA;  38520 MW;  FF7EB736962456F6 CRC64;

  Query Match             7.2%;  Score 82.5;  DB 1;  Length 352;
  Best Local Similarity   25.7%;  Pred. No. 6.9;
  Matches   43;  Conservative   27;  Mismatches   54;  Indels   43;  Gaps   10;

Qy         10 LKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAK 69
              |||:| |  |      |    :|  : |  : |||: : | :: |           || :|
Db        204 LKNIVACGAG-----FVDGLKLGDNTKAAVIRLGLMEMIRFVDV----------FYPGSK 248

Qy         70 NIAGFFH------LETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKE---GKHLGG 120
               :: ||       | |   | | :|:::   | :|  ::       || ||   |: | |
Db        249 -LSTFFESCGVADLITTCYG-GRNRRVSEAFVTSGKTIE-------DLEKEMLNGQKLQG 299

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSL------EVLPKSTRDAV 161
                 |  |: :: |    | ||  |    ::      :: ||   | :
Db        300 ---PPTAEEVN-YMLKNKGLEDKFPLFTAIHKICTNQLKPKDLIDCI 342


RESULT 12
PSGA_HUMAN
ID   PSGA_HUMAN     STANDARD;      PRT;   424 AA.
AC   Q15235;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Pregnancy-specific beta-1-glycoprotein 10 precursor (PSBG-10).
GN   PSG10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=91104939; PubMed=2271648;
RA   Barnett T.R., Pickle W. II, Elting J.J.;
RT   "Characterization of two new members of the pregnancy-specific beta
RT   1-glycoprotein family from the myeloid cell line KG-1 and suggestion
RT   of two distinct classes of transcription unit.";
RL   Biochemistry 29:10213-10218(1990).
CC   -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC   -!- DEVELOPMENTAL STAGE: PSBG ARE PRODUCED IN HIGH QUANTITY DURING
CC       PREGNANCY.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CEA
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; X17098; CAA34957.1; -.
DR   Genew; HGNC:9515; PSG10.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 4.
DR   SMART; SM00410; IG_like; 3.
DR   SMART; SM00408; IGc2; 1.
KW   Immunoglobulin domain; Glycoprotein; Signal; Repeat; Multigene family.
FT   SIGNAL        1     34       BY SIMILARITY.
FT   CHAIN        35    424       PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN
FT                                10.
FT   DOMAIN       35    143       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      161    223       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      254    316       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      346    400       IG-LIKE C2-TYPE DOMAIN 3.
FT   SITE        126    128       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    168    216       PROBABLE.
FT   DISULFID    261    309       PROBABLE.
FT   DISULFID    353    393       PROBABLE.
FT   CARBOHYD     61     61       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    198    198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267    267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  47676 MW;  0B7726090B31B92E CRC64;

  Query Match             7.2%;  Score 82.5;  DB 1;  Length 424;
  Best Local Similarity   22.3%;  Pred. No. 8.5;
  Matches   40;  Conservative   24;  Mismatches   60;  Indels   55;  Gaps    7;

Qy         56 ENEKALLNCYPNAKNIAGF----------FHLETDYVGLGIDRQM---------------- 90
              |  :  ||  :   :|: |:          :|  | ||   :| |:
Db         47 EGKDVLLLVHNLPQNLTGYIWYKGQMTDLYHYITSYV---VDGQIIYGPAYSGRETVYSN 103

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
              | | : |    ||||  |: :|| |   ||         :  |       | | ::: |
Db        104 ASLLIQNVTQEDAGS-YTLHIIKRGDGTGG---------VTGYFTVTLYSETPKRSISSS 153

Qy        151 EVLPKSTRDAVNY----------------GMVLSVIACIQHLAKNQKIYLCGGDAKYLS 193
               : |:   :||                |  | :   :|    |: :||  |  ||::
Db        154 NLNPREVMEAVRLICDPETPDASYLWLLNGQNLPMTHRLQLSKTNRTLYLF-GVTKYIA 211


RESULT 13
PL_HAEIN
ID   NPL_HAEIN      STANDARD;      PRT;   293 AA.
AC   P44539;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable N-acetylneuraminate lyase subunit (EC 4.1.3.3) (N-
DE   acetylneuraminic acid aldolase) (N-acetylneuraminate pyruvate lyase)
DE   (Nalase).
GN   NANA OR HI0142.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: N-acetylneuraminate = N-acetyl-D-mannosamine +
CC       pyruvate.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE DHDPS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32700; AAC21814.1; -.
DR   HSSP; P06995; 1NAL.
DR   TIGR; HI0142; -.
DR   InterPro; IPR002220; DHDPS.
DR   InterPro; IPR005264; NanA.
DR   Pfam; PF00701; DHDPS; 1.
DR   PRINTS; PR00146; DHPICSNTHASE.
DR   ProDom; PD001859; DHDPS; 1.
DR   TIGRFAMs; TIGR00683; nanA; 1.
DR   PROSITE; PS00665; DHDPS_1; 1.
DR   PROSITE; PS00666; DHDPS_2; 1.
KW   Lyase; Complete proteome.
FT   ACT_SITE    164    164       BY SIMILARITY.
SQ   SEQUENCE   293 AA;  32564 MW;  ACA755D0A5D93D33 CRC64;

  Query Match             7.1%;  Score 81.5;  DB 1;  Length 293;
  Best Local Similarity   20.8%;  Pred. No. 6.8;
  Matches   48;  Conservative   37;  Mismatches   79;  Indels   67;  Gaps   10;

Qy          6 SFTDLK---NLVLCDIGNTRIHFA--------QNYQLFSSAKEDLKRLGIQKEI--FYIS 52
              || ::|   : :: : ||  | ::          : |    :: | ||::     ||:
Db        114 SFPEIKHYYDTIIAETGNNMIVYSIPFLTGVNMGIEQFGELYKNPKVLGVKFTAGDFYLL 173

Qy         53 VNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLI 112
                : |   |||    |||  :      ||:|          | :   ||   :: :
Db        174 ------ERLKKAYPNHLIWAGFDEMMLPAASLGVD----------GAI---GSTFNVNGV 214

Qy        113 KEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIAC 172
              :              |: |    |: |: |       :||:        |   ::  ::|
Db        215 R-------------ARQIFELTKAGKLAE-------ALEI------QHVTNDLIEGILAN 248

Qy        173 IQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
               :|   : : | | || |          |:| :     |  | ||  |: |
Db        249 GLYLTIKELLKLEGVDAGY---------CREPMTSKATEEQLAKAKDLKAK 290


RESULT 14
ARCC_CLOPE
ID   ARCC_CLOPE     STANDARD;      PRT;   314 AA.
AC   Q46171;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Carbamate kinase (EC 2.7.2.2).
GN   ARCC OR CPE0171.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   MEDLINE=97199138; PubMed=9053381;
RA   Ohtani K., Bando M., Swe T., Banu S., Oe M., Hayashi H., Shimizu T.;
RT   "Collagenase gene (colA) is located in the 3'-flanking region of the
```

```
RT   perfringolysin O (pfoA) locus in Clostridium perfringens.";
RL   FEMS Microbiol. Lett. 146:155-159(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + NH(3) + CO(2) = ADP + carbamoyl
CC       phosphate.
CC   -!- PATHWAY: Arginine degradation via arginine deiminase; third step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CARBAMATE KINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X97768; CAA66367.1; -.
DR   EMBL; AP003185; BAB79877.1; -.
DR   HSSP; P95474; 1E19.
DR   InterPro; IPR001048; Aa_kinase.
DR   InterPro; IPR003964; Bac_carb_kinase.
DR   Pfam; PF00696; aakinase; 1.
DR   PRINTS; PR01469; CARBMTKINASE.
DR   TIGRFAMs; TIGR00746; arcC; 1.
KW   Transferase; Kinase; Arginine metabolism; Complete proteome.
FT   CONFLICT      9      9       G -> E (IN REF. 1).
FT   CONFLICT    126    126       K -> T (IN REF. 1).
SQ   SEQUENCE   314 AA;  33864 MW;  170C59630EBD1A37 CRC64;

  Query Match             7.1%;  Score 81.5;  DB 1;  Length 314;
  Best Local Similarity   23.2%;  Pred. No. 7.4;
  Matches   38;  Conservative   32;  Mismatches   55;  Indels   39;  Gaps    7;

Qy         29 YQLFSSAKEDLKRLGIQKEIFYIS---VNEENEKALLNCYPNAKNIAGFF------HLET 79
              | | :: :|:| : ||:| :  |:   : ::|:    |     | |  |:       ||
Db         90 YHLQNTIREELLKRGIEKSVDTITTQVIVDKNDPGFKN---PTKPIGSFYTKEEAEKLEK 146

Qy         80 DY-------VGLG-------------IDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLG 119
              |         | |             ::::     |::| :| |     | ::::|  |
          147 DKGYTMKEDAGRGYRRVVASPKPVDIVEKEAIKTMVDSGFIVIACGGGGIPVVEDGDRLE 206

Qy        120 GCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNY 163
              |  :| :     | :| |:||:      |:| :|    |  ||:
Db        207 G--VPAVIDKDFAAEKLAEILD-----ADALLILTAVDRVCVNF 243


RESULT 15
ACC8_HUMAN
ID   ACC8_HUMAN     STANDARD;      PRT;  1580 AA.
AC   Q09428; O75948; Q16583;
DT   01-NOV-1995 (Rel. 32, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sulfonylurea receptor 1.
GN   ABCC8 OR SUR1 OR SUR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Pancreatic islets;
RA   Gonzalez G., Aguilar-Bryan L., Bryan J.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Pancreas;
RA   Nishimura M., Miki T., Aizawa T., Seino S.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Foreskin, and Brain;
RA   Thomas P.T., Wohllk N., Huang E., Gagel R.F., Cote G.J.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1186-1580 FROM N.A.
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=95232533; PubMed=7716548;
RA   Thomas P.M., Cote G.J., Wohllk N., Haddad B., Mathew P.M., Rabl W.,
RA   Aguilar-Bryan L., Gagel R.F., Bryan J.;
RT   "Mutations in the sulfonylurea receptor gene in familial persistent
RT   hyperinsulinemic hypoglycemia of infancy.";
RL   Science 268:426-429(1995).
RN   [5]
RP   TOPOLOGY.
RX   MEDLINE=99436117; PubMed=10506167;
RA   Raab-Graham K.F., Cirilo L.J., Boettcher A.A., Radeke C.M.,
RA   Vandenberg C.A.;
RT   "Membrane topology of the amino-terminal region of the sulfonylurea
RT   receptor.";
RL   J. Biol. Chem. 274:29122-29129(1999).
RN   [6]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=99268411; PubMed=10338089;
RA   Meissner T., Beinbrech B., Mayatepek E.;
RT   "Congenital hyperinsulinism: molecular basis of a heterogeneous
RT   disease.";
RL   Hum. Mutat. 13:351-361(1999).
RN   [7]
RP   VARIANT PHHI VAL-715.
RX   MEDLINE=96354544; PubMed=8751851;
RA   Thomas P.M., Wohllk N., Huang E., Kuhnle U., Rabl W., Gagel R.F.,
RA   Cote G.J.;
RT   "Inactivation of the first nucleotide-binding fold of the
RT   sulfonylurea receptor, and familial persistent hyperinsulinemic
RT   hypoglycemia of infancy.";
RL   Am. J. Hum. Genet. 59:510-518(1996).
RN   [8]
RP   VARIANT ALA-1368.
RX   MEDLINE=96220088; PubMed=8635661;
RA   Inoue H., Ferrer J., Welling C.M., Elbein S.C., Hoffman M.,
RA   Mayorga R., Warren-Perry M., Zhang Y., Millns H., Turner R.,
RA   Province M., Bryan J., Permutt M.A., Aguilar-Bryan L.;
RT   "Sequence variants in the sulfonylurea receptor (SUR) gene are
RT   associated with NIDDM in Caucasians.";
RL   Diabetes 45:825-831(1996).
RN   [9]
RP   VARIANT PHHI PHE-1386 DEL, AND VARIANTS GLY-1359; ALA-1368 AND
RP   ILE-1571.
RX   MEDLINE=97081768; PubMed=8923011;
RA   Nestorowicz A., Wilson B.A., Schoor K.P., Inoue H., Glaser B.,
RA   Landau H., Stanley C.A., Thornton P.S., Clement J.P. IV, Bryan J.,
RA   Aguilar-Bryan L., Permutt M.A.;
RT   "Mutations in the sulfonylurea receptor gene are associated with
RT   familial hyperinsulinism in Ashkenazi Jews.";
RL   Hum. Mol. Genet. 5:1813-1822(1996).
RN   [10]
RP   VARIANT PHHI ARG-1477.
RX   MEDLINE=96256476; PubMed=8650576;
RA   Nichols C.G., Shyng S.-L., Nestorowicz A., Glaser B., Clement J.P. IV,
RA   Gonzalez G., Aguilar-Bryan L., Permutt M.A., Bryan J.;
RT   "Adenosine diphosphate as an intracellular regulator of insulin
RT   secretion.";
RL   Science 272:1785-1787(1996).
RN   [11]
RP   VARIANTS GLN-274; MET-559; ASN-809; CYS-833 AND ALA-1368.
```

```
RX   MEDLINE=98178572; PubMed=9519757;
RA   Ohta Y., Tanizawa Y., Inoue H., Hosaka T., Ueda K., Matsutani A.,
RA   Repunte V.P., Yamada M., Kurachi Y., Bryan J., Aguilar-Bryan L.,
RA   Permutt M.A., Oka Y.;
RT   "Identification and functional analysis of sulfonylurea receptor 1
RT   variants in Japanese patients with NIDDM.";
RL   Diabetes 47:476-481(1998).
RN   [12]
RP   VARIANTS PHHI.
RX   MEDLINE=98282238; PubMed=9618169;
RA   Nestorowicz A., Glaser B., Wilson B.A., Shyng S.-L., Nichols C.G.,
RA   Stanley C.A., Thornton P.S., Permutt M.A.;
RT   "Genetic heterogeneity in familial hyperinsulinism.";
RL   Hum. Mol. Genet. 7:1119-1128(1998).
RN   [13]
RP   VARIANTS PHHI PRO-1351; CYS-1419 AND TRP-1492.
RX   MEDLINE=98443218; PubMed=9769320;
RA   Verkarre V., Fournet J.-C., de Lonlay P., Gross-Morand M.-S.,
RA   Devillers M., Rahier J., Brunelle F., Robert J.-J., Nihoul-Fekete C.,
RA   Saudubray J.-M., Junien C.;
RT   "Paternal mutation of the sulfonylurea receptor (SUR1) gene and
RT   maternal loss of 11p15 imprinted genes lead to persistent
RT   hyperinsulinism in focal adenomatous hyperplasia.";
RL   J. Clin. Invest. 102:1286-1291(1998).
RN   [14]
RP   VARIANT PHHI ASP-186.
RX   MEDLINE=99265491; PubMed=10334322;
RA   Otonkoski T., Aemmaelae C., Huopio H., Cote G.J., Chapman J.,
RA   Cosgrove K., Ashfield R., Huang E., Komulainen J., Ashcroft F.M.,
RA   Dunne M.J., Kere J., Thomas P.M.;
RT   "A point mutation inactivating the sulfonylurea receptor causes the
RT   severe form of persistent hyperinsulinemic hypoglycemia of infancy in
RT   Finland.";
RL   Diabetes 48:408-415(1999).
CC   -!- FUNCTION: PUTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC       POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC       AND INSULIN RELEASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- DISEASE: DEFECTS IN ABCC8 ARE A CAUSE OF FAMILIAL PERSISTENT
CC       HYPERINSULINEMIC HYPOGLYCEMIA OF INFANCY (PHHI), AN AUTOSOMAL
CC       RECESSIVE DISORDER CHARACTERIZED BY UNREGULATED AND HIGH INSULIN
CC       SECRETION. IT CAUSES NESIDIOBLASTOSIS, A DIFFUSE ABNORMALITY OF
CC       THE PANCREAS IN WHICH THERE IS EXTENSIVE, OFTEN DISORGANIZED
CC       FORMATION OF NEW ISLETS. PHHI HAS A VARIABLE CLINICAL PHENOTYPE
CC       BUT USUALLY PRESENTS DURING THE NEONATAL PERIOD OR INFANCY WITH
CC       SEIZURES, COMA, AND OFTEN LARGE BIRTH WEIGHT FOR GESTATIONAL AGE.
CC   -!- DISEASE: DEFECTS IN ABCC8 MAY CONTRIBUTE TO AUTOSOMAL DOMINANT
CC       NON-INSULIN-DEPENDENT DIABETES MELLITUS TYPE II (MODY2 OR MODY-II)
CC       (ALSO KNOWN AS NIDDM), WHICH IS CHARACTERIZED BY AN AUTOSOMAL
CC       DOMINANT MODE OF INHERITANCE, ONSET DURING CHILDHOOD (USUALLY
CC       BEFORE 25 YEARS OF AGE) AND A PRIMARY DEFECT IN INSULIN SECRETION.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78243; AAB02417.1; -.
DR   EMBL; L78208; AAB02417.1; JOINED.
DR   EMBL; L78209; AAB02417.1; JOINED.
DR   EMBL; L78210; AAB02417.1; JOINED.
DR   EMBL; L78211; AAB02417.1; JOINED.
DR   EMBL; L78212; AAB02417.1; JOINED.
DR   EMBL; L78255; AAB02417.1; JOINED.
DR   EMBL; L78213; AAB02417.1; JOINED.
DR   EMBL; L78214; AAB02417.1; JOINED.
DR   EMBL; L78215; AAB02417.1; JOINED.
DR   EMBL; L78216; AAB02417.1; JOINED.
DR   EMBL; L78217; AAB02417.1; JOINED.
DR   EMBL; L78218; AAB02417.1; JOINED.
DR   EMBL; L78219; AAB02417.1; JOINED.
DR   EMBL; L78220; AAB02417.1; JOINED.
DR   EMBL; L78221; AAB02417.1; JOINED.
DR   EMBL; L78222; AAB02417.1; JOINED.
DR   EMBL; L78223; AAB02417.1; JOINED.
DR   EMBL; L78225; AAB02417.1; JOINED.
DR   EMBL; L78254; AAB02417.1; JOINED.
DR   EMBL; L78226; AAB02417.1; JOINED.
DR   EMBL; L78227; AAB02417.1; JOINED.
DR   EMBL; L78228; AAB02417.1; JOINED.
DR   EMBL; L78229; AAB02417.1; JOINED.
DR   EMBL; L78230; AAB02417.1; JOINED.
DR   EMBL; L78231; AAB02417.1; JOINED.
DR   EMBL; L78232; AAB02417.1; JOINED.
DR   EMBL; L78233; AAB02417.1; JOINED.
DR   EMBL; L78234; AAB02417.1; JOINED.
DR   EMBL; L78235; AAB02417.1; JOINED.
DR   EMBL; L78236; AAB02417.1; JOINED.
DR   EMBL; L78237; AAB02417.1; JOINED.
DR   EMBL; L78238; AAB02417.1; JOINED.
DR   EMBL; L78239; AAB02417.1; JOINED.
DR   EMBL; L78240; AAB02417.1; JOINED.
DR   EMBL; L78241; AAB02417.1; JOINED.
DR   EMBL; L78242; AAB02417.1; JOINED.
DR   EMBL; L78243; AAB02418.1; -.
DR   EMBL; L78208; AAB02418.1; JOINED.
DR   EMBL; L78209; AAB02418.1; JOINED.
DR   EMBL; L78210; AAB02418.1; JOINED.
DR   EMBL; L78211; AAB02418.1; JOINED.
DR   EMBL; L78212; AAB02418.1; JOINED.
DR   EMBL; L78255; AAB02418.1; JOINED.
DR   EMBL; L78213; AAB02418.1; JOINED.
DR   EMBL; L78214; AAB02418.1; JOINED.
DR   EMBL; L78215; AAB02418.1; JOINED.
DR   EMBL; L78216; AAB02418.1; JOINED.
DR   EMBL; L78217; AAB02418.1; JOINED.

  Query Match             7.0%;  Score 81;  DB 1;  Length 1580;
  Best Local Similarity   21.1%;  Pred. No. 51;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                | ::
Db       1174 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1231

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |          :  :|:|  :    |  :::|        |:|   |  :|  |:
Db       1232 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1288

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                 ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1289 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1345

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1346 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1403

Qy        208 DGMEIA 213
               ||::||
Db       1404 DGIDIA 1409


Search completed: June 24, 2003, 21:51:24
Job time : 5.87859 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 8.13374 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-14
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	2	F64627	hypothetical prote
2	1115	96.7	223	2	G71887	hypothetical prote
3	391.5	34.0	209	2	H81382	hypothetical prote
4	165	14.3	246	2	D72320	conserved hypothet
5	159.5	13.8	248	2	H83111	hypothetical prote
6	158	13.7	276	2	AI2292	hypothetical prote
7	141.5	12.3	257	2	S75559	hypothetical prote
8	138.5	12.0	229	2	E70465	hypothetical prote
9	117.5	10.2	273	2	E97293	probable transcrip
10	116.5	10.1	592	2	B81009	BirA protein/Bvg a
11	116.5	10.1	592	2	H82031	probable biotin-[a
12	116	10.1	224	2	A99571	conserved hypothet
13	115.5	10.0	242	2	A82637	conserved hypothet
14	109.5	9.5	233	2	S66100	conserved hypothet
15	94	8.2	407	2	F69805	hypothetical prote
16	93.5	8.1	1164	2	C86238	protein T10O24.5 [
17	91	7.9	254	2	F83660	hypothetical prote
18	90.5	7.8	274	2	H86937	conserved hypothet
19	90.5	7.8	482	2	F72278	conserved hypothet
20	89.5	7.8	232	2	H69315	cytochrome-c3 hydr
21	89.5	7.8	272	2	A70955	hypothetical prote
22	89.5	7.8	1582	2	A56248	sulfonylurea recep
23	89	7.7	243	2	T30671	hypothetical prote
24	88.5	7.7	262	2	F70165	conserved hypothet
25	88.5	7.7	1182	2	D71638	DNA polymerase III
26	87.5	7.6	344	2	A64103	DNA-directed DNA p
27	87.5	7.6	483	2	T03204	hypothetical prote
28	86.5	7.5	393	2	AD2219	hypothetical prote
29	86	7.5	474	2	B82923	pyruvate kinase UU
30	86	7.5	627	2	AH2851	hypothetical prote
31	86	7.5	702	2	F97628	ATP-binding protei
32	85.5	7.4	240	2	AI0941	probable GntR-fami
33	85.5	7.4	401	2	E83720	methyl-accepting c
34	85.5	7.4	426	2	C64086	glycerol-3-phospha
35	84	7.3	764	2	F71863	phenylalanine-tRNA
36	83.5	7.2	450	2	AC1079	PTS system, cellob
37	83.5	7.2	452	2	AB1437	PTS system, cellob
38	83.5	7.2	839	2	B96998	phosphoenolpyruvat
39	83	7.2	303	1	D64070	ATP phosphoribosyl
40	83	7.2	553	2	T02475	hypothetical prote
41	82.5	7.2	259	2	AF1102	conserved hypothet
42	82.5	7.2	265	2	T36391	hypothetical prote
43	82.5	7.2	352	2	JS0023	glycerol-3-phospha
44	82.5	7.2	353	2	S31790	glycerol-3-phospha
45	82.5	7.2	363	2	S23137	glycerol-3-phospha

ALIGNMENTS

RESULT 1
F64627
hypothetical protein HP0862 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
C;Accession: F64627
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64627
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <TOM>
A;Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07916.1; PID:g23
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 100.0%; Score 1153; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 7.5e-93;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 2
G71887
hypothetical protein jhp0796 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Aug-2000
C;Accession: G71887
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric

```
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: G71887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <ARN>
A;Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g415536
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0796
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

  Query Match             96.7%;  Score 1115;  DB 2;  Length 223;
  Best Local Similarity   94.2%;  Pred. No. 1.5e-89;
  Matches  210;  Conservative   11;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||| |||:|:||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFKDLKDLILCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||:|||||||||||| |||:|||||||||||:||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYIGLGIDRQMACLAVVNGVIVDAGSAITIDLVKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              |||||||||:|||||||||||||||||||||||:|||||||||:||:|:|||||||:|
Db        121 CILPGLAQYVHAYKKSAKILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCIQHLAKDQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223


RESULT 3
H81382
hypothetical protein Cj0394c [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: H81382
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: H81382
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74230.1; PID:g696786
A;Experimental source: serotype O2, strain NCTC 11168
  Genetics:
A;Gene: Cj0394c
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

  Query Match             34.0%;  Score 391.5;  DB 2;  Length 209;
  Best Local Similarity   39.2%;  Pred. No. 1.1e-26;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::||||||:  :|  : : |:   :        :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203


RESULT 4
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             14.3%;  Score 165;  DB 2;  Length 246;
  Best Local Similarity   23.2%;  Pred. No. 6.6e-07;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:          |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 -----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                        :: |:||   |  :|      |   | || :           :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|       ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

Qy        144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
              | |   |: | | : :  |:|         :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
              | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 5
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; I
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

  Query Match             13.8%;  Score 159.5;  DB 2;  Length 248;
  Best Local Similarity   28.6%;  Pred. No. 2e-06;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;
```

```
Qy          51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
               :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| :: 
Db          53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy          98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                 :|:| |:|:| ||:  :| |||| | ||:              :|  |  : | 
Db         113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy         140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
                 || :|          | :||  | :| :         :|| : |  : :|:| ||||: 
Db         173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy         191 YLSAFLPHSVCKERLVFDGMEIA 213
               :   |  :     ||| |: :| 
Db         222 LVRDELAGARIMPDLVFVGLALA 244
```

RESULT 6
AI2292
ypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

 Query Match 13.7%; Score 158; DB 2; Length 276;
 Best Local Similarity 25.1%; Pred. No. 3.1e-06;
 Matches 70; Conservative 51; Mismatches 94; Indels 64; Gaps 13;

```
Qy           1 MPARQSFTDLKNLVLC-DIGNTRIHFA-----------------QNYQLFSSAKEDLKR 41
               | ::||   ::|  |  :|||:|:|:|                 |   : :  |: 
Db           1 MISQQSAKHIENQWLALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEV 60

Qy          42 LGIQKEIF---------------YISVNEENEKALLNCYPNAKNIA-GFFHLETDYVGLG 85
                  :|||                :|:     :  |   | | : |      |   |  ||
 b          61 GSEEKEIFFTFFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPTLG 120

Qy          86 IDRQMACLAVNNG-----VVVDAGSAITIDLIKEGKHL-GGCILPGLA-QYIHAYKKSAK 138
               |||  :|             :|:|||:|:|      ||:| || ||||:  |:    ::: : 
Db         121 IDRALALWGAGMSWGFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQ 180

Qy         139 ILEQPFKALDSLEVLPK---STRDAVNYGMVLSVIACIQ-------HLAKNQKIYLCGGD 188
               : :   :|: ||   |:   :| :|:  |:: ::|| ::        |  : |: : |||
Db         181 LPQVEMEAIKSLP--PRFALNTTEAIQSGVIYTLIAGMRDFTEEWLSLFPDGKVAIKGGD 238

Qy         189 ----AKYLSAFLP----HSVCKERLVFDGMEIALKKAGI 219
                   || |  |      : :  |:| ||:  :  ||:
Db         239 RILLLNYLQALYPDLAARLIVEPNLIFWGMQTIV--AGV 275
```

RESULT 7
S75559
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75559
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocy
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75559
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-257 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18120.1; PID:d10
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

 Query Match 12.3%; Score 141.5; DB 2; Length 257;
 Best Local Similarity 26.9%; Pred. No. 7.6e-05;
 Matches 66; Conservative 28; Mismatches 76; Indels 75; Gaps 12;

```
Qy          18 IGNTRIHFAQNYQLFSSAKEDLKR---------------LGIQKEIFYISVNEENE---- 58
               |||:|:|:|     : |    |:                 ||    :    | |: | 
Db          25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy          59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                         |:| ||:                 |||| :| |          :|||
Db          80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy         104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
               |:|:||    : | | || |||||   :       ::   |   :| |  ||      | 
Db         124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy         159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
                 |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||        : | 
Db         182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

Qy         206 VFDGM 210
               :| ||
Db         242 IFLGM 246
```

RESULT 8
E70465
hypothetical protein aq_1924 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: E70465
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-229 <AQF>
A;Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AE0
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1924

 Query Match 12.0%; Score 138.5; DB 2; Length 229;
 Best Local Similarity 26.5%; Pred. No. 0.00012;
 Matches 61; Conservative 32; Mismatches 84; Indels 53; Gaps 11;

```
Qy          17 DIGNTRIHFA-------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
               |:||: :  |              : :       || ||       ||| :   : : 
Db           7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy          64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                 |  | :    | :: ||     || ||      :       | ||: |||:|: |||: | 
Db          60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy         115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
               ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
Db         120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172
```

```
Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                  : | :   : : | : |: : ||: || | |    :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219
```

RESULT 9
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
:Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

 Query Match 10.2%; Score 117.5; DB 2; Length 273;
 Best Local Similarity 21.4%; Pred. No. 0.01;
 Matches 57; Conservative 48; Mismatches 94; Indels 67; Gaps 11;

```
Qy         10 LKNLVLCDIGNTRI--------HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKAL 61
              || ::: |:||| |             ::| :         |||  :  :   :: :  |
Db         10 LKVILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQ--VMNLFQQDKLDPTL 67

Qy         62 LN-----------CYPNAKNIAGFFHLETDYVGLGI---------------DRQMACLA 94
              :             |     |  :| :    || ||                || :  :|
Db         68 VEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVA 127

Qy         95 VN-----NGVVVDAGSAITIDLIKE-GKHLGGCILPGLAQYIHA-YKKSAKI----LEQP 143
               :     : :::| |:| |   ::| | :|| | ||:     | ::|:||:    | :|
Db        128 AHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKP 187

Qy        144 FKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQK-----------IYLCGGDAKY 191
                |      : |:|  ::  |:|   |  :::: :  |            :   || ||
Db        188 AYA------ICKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKL 241

Qy        192 LSAFLPH-SVCKERLVFDGMEIALKK 216
              :|    :  |    |  :|: |  :|
Db        242 ISEEAKNVDVINPFLTLEGLRIIYEK 267
```

RESULT 10
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain MC
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

 Query Match 10.1%; Score 116.5; DB 2; Length 592;
 Best Local Similarity 24.9%; Pred. No. 0.032;
 Matches 64; Conservative 37; Mismatches 85; Indels 71; Gaps 13;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
                :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                 : |  |||  |:|:|:| |  :| :||| |:||    |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||          | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567
```

RESULT 11
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitid
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C;Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z249
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

 Query Match 10.1%; Score 116.5; DB 2; Length 592;
 Best Local Similarity 24.9%; Pred. No. 0.032;
 Matches 64; Conservative 37; Mismatches 85; Indels 71; Gaps 13;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
                :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                 : |  |||  |:|:|:| |  :| :||| |:||    |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||          | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567
```

RESULT 12

```
A99571
conserved hypothetical protein. MYPU_4730 [imported] - Mycoplasma pulmonis (strain UAB C
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: A99571
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: A99571
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-224 <KUR>
A;Cross-references: GB:AL445566; PID:g14089887; PIDN:CAC13646.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_4730
A;Genetic code: SGC3

  Query Match           10.1%;  Score 116;  DB 2;  Length 224;
  Best Local Similarity  22.3%;  Pred. No. 0.011;
  Matches   51;  Conservative   30;  Mismatches   94;  Indels   54;  Gaps    7;

Qy         14 VLCDIGNTRIHFA-----------------------QNYQLFSSAKEDLKRLGIQKEI 48
              :  |:||:   ||                       ::: ||   | |:      |||
Db          3 IFIDLGNSFAKFALKNENKAHFLFRLKTQSVVDPSFEIKSFNLFEFNKLDV------KEI 56

Qy         49 FYISVNEENEKALL-----NCYPNAKNIAGFFHLE----------TDYVGLGI--DRQM 90
                ||    |  :|     : : ||| |  | |            |  :|| |  :
Db         57 LICSVRNAKENQILEFKLKSIFKNAK-IDFFIHKSQSLVKFCQKELTSEIGLDIVANAYY 115

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
                | :  |:|| |  ||     |  | ||: |    :      :|  |  :
Db        116 VLHKSKNAIFISLGTAIVITQIKNSSIEGISIYPGIYQSFKNFFNVVAKIESNFNIIKIP 175

Qy        151 EVLPKSTRDAVN----YGMVLSVIACIQHLAKNQKIYLCGGDAKYLSAF 195
               :| |:| ::::     | |  :   |  : |   |:: ||  :   |
Db        176 PILGKTTLESISSSLVRGSVFLLKGYIDEIDKTSDIFITGGTITFSKIF 224


RESULT 13
A82637
conserved hypothetical protein XF1795 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84603.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1795

  Query Match           10.0%;  Score 115.5;  DB 2;  Length 242;
  Best Local Similarity  31.5%;  Pred. No. 0.013;
  Matches   45;  Conservative   22;  Mismatches   55;  Indels   21;  Gaps    6;

Qy         85 GIDRQMACL---AVNNGVVVDAGSAITIDLI-KEGKHLGGCILPG---LAQYIHAYKKSA 137
              |:|| :| |      | :||  |:|:||||:   | |||| |      : | :|     |
Db         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152

Qy        138 KILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACI--------QHLAKNQKIYLCGGDA 189
              :  : |    : ||   : | ||:  |   : :| |        | | :: :: | ||
Db        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211

Qy        190 KYLSAFLPHSVCKERLVFDGMEI 212
                |  :|   | :  || ||: |
Db        212 ASLLPWLGDVVHRPTLVLDGLAI 234


RESULT 14
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S66100; E69740
R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis c
A;Reference number: S65967; MUID:96051385; PMID:7584024
A;Accession: S66100
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <OGA>
A;Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g46
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 199
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e1182
A;Experimental source: strain 168
C;Genetics:
A;Gene: yacB
A;Start codon: TTG
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match           9.5%;  Score 109.5;  DB 2;  Length 233;
  Best Local Similarity  26.7%;  Pred. No. 0.041;
  Matches   35;  Conservative   24;  Mismatches   47;  Indels   25;  Gaps    5;

Qy         74 FFHLETDYVG----------------LGIDRQMACLAV-----NNGVVVDAGSAITIDLI 112
              :||:|   ||                :| || :  :|       |  :||| |:| |   |
Db         80 YFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYI 139

Qy        113 KEGK-HLGGCILPGLAQYIHA-YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVI 170
               | | ::|| | ||:     | | ::||:        |:  :: |:|  |:  |::   :
Db        140 DENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDN--IIGKNTVSAMQSGILFGYV 197
```

```
Qy        171 ACIQHLAKNQK 181
                :: : |  |
Db        198 GQVEGIVKRMK 208


RESULT 15
F69805
hypothetical protein yfjB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: F69805
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
.;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69805
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-407 <KUN>
A;Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12645.1; PID:g2633140
A;Experimental source: strain 168
C;Genetics:
A;Gene: yfjB
C;Superfamily: Bacillus subtilis hypothetical protein yfjB

  Query Match             8.2%;  Score 94;  DB 2;  Length 407;
  Best Local Similarity   21.8%;  Pred. No. 1.8;
  Matches   52;  Conservative   25;  Mismatches   70;  Indels   92;  Gaps    7;

Qy         15 LCDIGNTRIHF----------------------AQNYQLFSSAKEDLKRLGIQKEIFY 50
               | |:| | : |                      |:||  :    | |||
Db        213 LKDVGKTGLRFAKGNAITATLITGATETFGAGVKIAENYAKYQDKPEVLKR--------- 263

Qy         51 ISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITID 110
                 || ||: |   |   :||              |   :   :   ||   |:
Db        264 -----ENAKAVGNAVNNTIFVAG-------------GATAGAVIGGAIGSFAGPVGTVL 304

          ?       111 LIKEGKHLGGCILPGLAQYIHAY-KKSAKILEQPF--------KALDSLEVLPKSTRDAV 161
                    |   | ::|| :   :|:|   : :|:|  |::|         | |:|   : ||  | :
Db        305 LGAAGSYIGGVVGEQIAKYTAGFAEKAALKLKEPIHAVVDTAKKGLESAGKVVKSVNDGI 364

Qy        162 NYGMVLSVIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGIL 220
               :              | |: |:     ||                  ||| | | |  |
Db        365 D-------------AANKSIHKGIDSAK----------------KGMEKAKKTADSL 392


Search completed: June 24, 2003, 22:03:14
Job time : 10.1337 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 22.895 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-14
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	19	AAW98422	H. pylori GHPO 344
2	1153	100.0	223	23	AAU91161	Helicobacter pylor
3	1153	100.0	223	23	AAU91181	Pantothenate kinas
4	391.5	34.0	209	23	AAU91165	Pantothenate kinas
5	165	14.3	246	23	AAU91156	Thermotoga maritim
6	159.5	13.8	248	23	AAU91164	Pantothenate kinas
7	148	12.8	249	23	AAU91182	Pantothenate kinas
8	144.5	12.5	241	23	AAU91179	Pantothenate kinas
9	144.5	12.5	249	23	AAU91178	Pantothenate kinas
10	141.5	12.3	257	23	AAU91160	Synechocystis pant
11	138.5	12.0	229	23	AAU91159	Aquifex aeolicus p
12	129.5	11.2	244	23	AAU91168	Pantothenate kinas
13	116.5	10.1	455	21	AAY74910	Neisseria meningit
14	116.5	10.1	592	20	AAY38615	Neisseria meningit
15	116.5	10.1	592	20	AAY38616	Neisseria meningit
16	116.5	10.1	592	21	AAY74912	Neisseria meningit
17	116.5	10.1	592	21	AAY74913	Neisseria meningit
18	116.5	10.1	592	23	AAU91166	Pantothenate kinas
19	116.5	10.1	592	23	AAU91169	Pantothenate kinas
20	116	10.1	455	20	AAY38617	Neisseria gonorrho
21	116	10.1	455	21	AAY74908	Neisseria gonorrhe
22	116	10.1	460	23	AAU91167	Pantothenate kinas
23	116	10.1	592	20	AAY38618	Neisseria gonorrho
24	116	10.1	592	21	AAY74911	Neisseria gonorrhe
25	115.5	10.0	242	23	AAU91180	Pantothenate kinas
26	111.5	9.7	389	21	AAY74909	Neisseria meningit
27	109.5	9.5	233	23	AAU91163	Pantothenate kinas
28	109.5	9.5	258	22	AAU01243	B. subtilis novel
29	109.5	9.5	258	23	AAU91149	Bacillus subtilis
30	108	9.4	219	23	AAU91176	Pantothenate kinas
31	105.5	9.2	257	23	AAU91174	Pantothenate kinas
32	102.5	8.9	250	23	AAU91150	Clostridium acetob
33	102	8.8	212	23	AAU91177	Pantothenate kinas
34	101	8.8	256	23	AAU91175	Pantothenate kinas
35	100.5	8.7	255	23	AAU91154	Geobacter sulfurre
36	98.5	8.5	262	23	AAU91170	Pantothenate kinas
37	94.5	8.2	338	22	ABB67066	Drosophila melanog
38	91	7.9	254	23	AAU91171	Pantothenate kinas
39	89.5	7.8	272	22	AAG81225	Mycobacterium tube
40	89.5	7.8	272	23	AAU91152	Mycobacterium tube
41	89.5	7.8	1582	16	AAR77088	Hamster sulphonylu
42	89	7.7	594	22	ABB61659	Drosophila melanog
43	88.5	7.7	262	23	AAU91158	Borrelia burgdorfe
44	87.5	7.6	258	23	AAU91172	Pantothenate kinas
45	87.5	7.6	344	23	ABB84735	DNA polymerase III

ALIGNMENTS

RESULT 1
AAW98422
ID AAW98422 standard; Protein; 223 AA.
XX
AC AAW98422;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 344 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14141.
XX

PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 709-710; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 223 AA;

Query Match 100.0%; Score 1153; DB 19; Length 223;
Best Local Similarity 100.0%; Pred. No. 4.2e-111;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
Db 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
Db 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
Db 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
Db 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223

RESULT 2
AAU91161
ID AAU91161 standard; Protein; 223 AA.
XX
AC AAU91161;
XX
DT 05-JUN-2002 (first entry)
XX
DE Helicobacter pylori pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Helicobacter pylori.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 78-79; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 223 AA;

Query Match 100.0%; Score 1153; DB 23; Length 223;
Best Local Similarity 100.0%; Pred. No. 4.2e-111;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
Db 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
Db 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
Db 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
Db 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223

RESULT 3
AAU91181
ID AAU91181 standard; Protein; 223 AA.
XX
AC AAU91181;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #19.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Helicobacter pylori.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54202.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 112-113; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate

CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 223 AA;

Query Match 100.0%; Score 1153; DB 23; Length 223;
Best Local Similarity 100.0%; Pred. No. 4.2e-111;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 4
AAU91165
ID AAU91165 standard; Protein; 209 AA.
XX
AC AAU91165;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #3.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Campylobacter jejuni.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 84-85; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 209 AA;

Query Match 34.0%; Score 391.5; DB 23; Length 209;
Best Local Similarity 39.2%; Pred. No. 2.7e-32;
Matches 80; Conservative 44; Mismatches 79; Indels 1; Gaps 1;

```
Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::|||||||:  :|  : : |:   :       :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
              :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |   |   ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
```

RESULT 5
AAU91156
ID AAU91156 standard; Protein; 246 AA.
XX
AC AAU91156;
XX
DT 05-JUN-2002 (first entry)
XX
DE Thermotoga maritima pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Thermotoga maritima.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 74-75; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This

```
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   246 AA;

  Query Match           14.3%;  Score 165;  DB 23;  Length 246;
  Best Local Similarity  23.2%;  Pred. No. 9.7e-09;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:          |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 -----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                        :: |:||   |  :|     |   | || :           :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|       ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

          144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
               |  |   |: | | : :  |:|         :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
               | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 6
AAU91164
ID   AAU91164 standard; Protein; 248 AA.
XX
AC   AAU91164;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #2.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 83-84; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   248 AA;

  Query Match           13.8%;  Score 159.5;  DB 23;  Length 248;
  Best Local Similarity  28.6%;  Pred. No. 3.7e-08;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
              :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::
Db         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                  :|:| |:|:| ||:  :| |||| | ||:              :|  |  : |
Db        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
               || :|          | :||  | :| :          :|| : |  : :|:| ||||:
Db        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy        191 YLSAFLPHSVCKERLVFDGMEIA 213
               :   |  :     ||| |: :|
Db        222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 7
AAU91182
ID   AAU91182 standard; Protein; 249 AA.
XX
AC   AAU91182;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #20.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas syringae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54203.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 113-114; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
```

```
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match             12.8%;  Score 148;  DB 23;  Length 249;
  Best Local Similarity   29.3%;  Pred. No. 5.7e-07;
  Matches   58;  Conservative   34;  Mismatches   78;  Indels   28;  Gaps    7;

Qy         34 SAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACL 93
              ||:|  | :    : | ||           |   |  :||  :   |:  ||:|| :| :
Db         58 SAEETAKLVCALADTFSISP---------VCAAPAPELAGVINGYDDFARLGLDRWLAFV 108

Qy         94 AVNNGV-----VVDAGSAITIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILE----QP 143
                : |     |:| |:||| | ::  |||||| | ||:    :  :   : :     :
Db        109 GAYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEA 168

Qy        144 FKALDSLEVLPKSTRDAVNYG--MVLSVIACIQ------HLAKNQKIYLCGGDAKYLSAF 195
               |||  | |  ::| :||  |  ::|   |  |      :   :  |:: |||| ::
Db        169 EKALVRL-VPGRATAEAVERGCSLMLRGFAMTQIEIAREYWGDDFAIFVTGGDAVLVADV 227

Qy        196 LPHSVCKERLVFDGMEIA 213
              || :     ||| |: :|
Db        228 LPGARIVPDLVFVGLALA 245


RESULT 8
AAU91179
ID   AAU91179 standard; Protein; 241 AA.
XX
AC   AAU91179;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #17.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Thiobacillus ferrooxidans.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54200.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 109-110; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   241 AA;

  Query Match             12.5%;  Score 144.5;  DB 23;  Length 241;
  Best Local Similarity   24.4%;  Pred. No. 1.3e-06;
  Matches   60;  Conservative   40;  Mismatches   97;  Indels   49;  Gaps   10;

Qy         13 LVLCDIGNTRIHFAQNYQ--LFSSAK--------EDLKRLGIQKEIFYISVNEENEKALL 62
              ::   :||||   |  :    | |          | |:: |:     ::|       ||
Db          1 MIFIAVGNTRTLLAHTHDGVHFDSVSVATSLPPTEILQQPGLT----WLSAPNREPVALG 56

Qy         63 NCYPNA--------------KNIAGFFH--LETDY---VGLGIDRQMACLAV------NN 97
                | |               :   |||   :  ||     || ||:   ||        :
Db         57 GVVPAALAAWREALATAEVREPDPGFFRRAVPHDYHPPESLGFDRRCCLLAAAMDYPGQD 116

Qy         98 GVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDS-LEVLPKS 156
              :|:| |:||||||:  |   || ||||:|  :    :  :|  |   |::  |:|
Db        117 SIVIDMGTAITIDLLAGGHFRGGRILPGIAMSLRGLHEGTALL--PEVVLNAPAEMLGND 174

Qy        157 TRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDG 209
              | :|:  |::          |   :  :  :| : ||||:     :  |: :  |:  |
Db        175 TSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRG 234

Qy        210 MEIALK 215
                : ::
Db        235 FYLWIR 240


RESULT 9
AAU91178
ID   AAU91178 standard; Protein; 249 AA.
XX
AC   AAU91178;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #16.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas putida.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54199.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 108-109; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
```

```
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match             12.5%;  Score 144.5;  DB 23;  Length 249;
  Best Local Similarity   29.8%;  Pred. No. 1.3e-06;
  Matches   61;  Conservative   26;  Mismatches   63;  Indels   55;  Gaps   10;

Qy         51 ISVNEENEKALLNC--------------YPNAKNIAGFFHLETDYVGLGIDRQMAC---- 92
              :||  | |   | |               :|  : :||  :   ||  ||:|| :|
Db         54 VSVRSEEETDAL-CALIAQAFAVQAKVAHP-VREMAGVRNGYDDYQRLGMDRWLAALGAF 111

Qy         93 -LAVNNGVVVDAGSAITIDLIK-EGKHLGGCILPGL----------AQYIHAYKKSAKIL 140
               ||    :|:| |:|   | :  :|:|||| | ||:          : |     ||:
Db        112 HLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRIRYDDASAE-- 169

Qy        141 EQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ------------KIYLCGGD 188
              :|| ||    :|| :||  | ||     :|  |  |              ::| |||
Db        170 ----RALSSLSP-GRSTVEAVERGCVL----MLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        189 AKYLSAFLPHSVCKERLVFDGMEIA 213
              |  : | || :     ||| |: :|
Db        221 APLVRAALPQARVVPDLVFVGLAMA 245


RESULT 10
AAU91160
ID   AAU91160 standard; Protein; 257 AA.
XX
AC   AAU91160;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Synechocystis pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Synechocystis sp.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 77-78; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   257 AA;

  Query Match             12.3%;  Score 141.5;  DB 23;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 2.8e-06;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy         18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
              |||:|:|:|     : |    |:                  ||    :    | |: |
Db         25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy         59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                        |:| ||:                  |||| :| |          :|||
Db         80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy        104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
              |:|:||    : | | || |||||   :        ::    |   :| |  ||      |
Db        124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy        159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
               |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||        : |
Db        182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

Qy        206 VFDGM 210
              :| ||
Db        242 IFLGM 246


RESULT 11
AAU91159
ID   AAU91159 standard; Protein; 229 AA.
XX
AC   AAU91159;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Aquifex aeolicus pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Aquifex aeolicus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 77; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
```

```
SQ   Sequence   229 AA;

  Query Match            12.0%;  Score 138.5;  DB 23;  Length 229;
  Best Local Similarity  26.5%;  Pred. No. 4.9e-06;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy         17 DIGNTRIHFA------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
              |:||: :  |              : :      || ||       ||| :   : :
Db          7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy         64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                 |  | :    | :: ||     || ||      :       | ||: ||:|: |||: |
Db         60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy        115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
              ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
Db        120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172

Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                 : | :   : : | : |: : ||: || | |    :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219


RESULT 12
AAU91168
ID   AAU91168 standard; Protein; 244 AA.
XX
AC   AAU91168;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #6.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
X
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54189.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 8; Page 94-95; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   244 AA;

  Query Match            11.2%;  Score 129.5;  DB 23;  Length 244;
  Best Local Similarity  26.0%;  Pred. No. 4.6e-05;
  Matches   58;  Conservative   40;  Mismatches   82;  Indels   43;  Gaps   12;

Qy         12 NLVLCDIGNT--RIHFAQNYQLFS------SAKEDLKRL----GIQKEIF-YISVNEENE 58
              ||:: | ||:  :: | :|  : |       | : |  |       | |:  : : :|
Db          4 NLIV-DQGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEA 62

Qy         59 KALLNCYPNAKNIAGF-----FHLETDYVGLGIDRQMACLAVNN------GVVVDAGSAI 107
               :|::     |  : |       |: |   || ||  | :  ::       :|:|||:||
Db         63 EAIVRSCAAASLMMGTETPVPLRLQYDRRTLGADRLAAVVGAHSLYPNTELLVIDAGTAI 122

Qy        108 TIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK------STRDA 160
              | : :  || :||| | |||    |   |:  :       :|   : ||      || :|
Db        123 TYERVSAEGIYLGGNISPGL----HLRFKALHLFTGRLPLIDPSGISPKIAEYGSSTEEA 178

Qy        161 VNYGMVLSVIACIQ------HLAKNQK-IYLCGGDAKYLSAFL 196
              :  |::  :   |       |  : :  : | |||| ||:  :
Db        179 ITAGVIHGLAGEIDRYIDDLHAKEGRSAVILTGGDANYLARII 221


RESULT 13
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
```

```
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             10.1%;  Score 116.5;  DB 21;  Length 455;
  Best Local Similarity   24.9%;  Pred. No. 0.0024;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        194 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 250

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |       | ||    |
Db        251 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 306

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        307 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 362

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||        | :| :||  ||: :|   :  :           |   : : || |
Db        363 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 413

Qy        190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        414 AKVAEALPPAFLAENTV 430


RESULT 14
AAY38615
ID   AAY38615 standard; Protein; 592 AA.
XX
AC   AAY38615;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             10.1%;  Score 116.5;  DB 20;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0034;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |       | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||        | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 15
AAY38616
ID   AAY38616 standard; Protein; 592 AA.
XX
AC   AAY38616;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
```

```
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             10.1%;  Score 116.5;  DB 20;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0034;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |       | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                 : |  |||  |:|:|:|:| |  :| :|||:|:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
               ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


Search completed: June 24, 2003, 21:46:21
Job time : 23.895 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 12.3931 Seconds
(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453A-67
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	9	US-09-895-913A-74	Sequence 74, Appl
2	1153	100.0	223	9	US-09-813-453A-14	Sequence 14, Appl
3	1153	100.0	223	9	US-09-813-453A-67	Sequence 67, Appl
4	391.5	34.0	209	9	US-09-813-453A-21	Sequence 21, Appl
5	165	14.3	246	9	US-09-813-453A-9	Sequence 9, Appli
6	159.5	13.8	248	9	US-09-813-453A-20	Sequence 20, Appl
7	148	12.8	249	9	US-09-813-453A-70	Sequence 70, Appl
8	144.5	12.5	241	9	US-09-813-453A-63	Sequence 63, Appl
9	144.5	12.5	249	9	US-09-813-453A-61	Sequence 61, Appl
10	141.5	12.3	257	9	US-09-813-453A-13	Sequence 13, Appl
11	138.5	12.0	229	9	US-09-813-453A-12	Sequence 12, Appl
12	129.5	11.2	244	9	US-09-813-453A-41	Sequence 41, Appl
13	116.5	10.1	592	9	US-09-813-453A-22	Sequence 22, Appl
14	116.5	10.1	592	9	US-09-813-453A-43	Sequence 43, Appl
15	116	10.1	460	9	US-09-813-453A-39	Sequence 39, Appl
16	115.5	10.0	242	9	US-09-813-453A-65	Sequence 65, Appl
17	109.5	9.5	233	9	US-09-813-453A-17	Sequence 17, Appl
18	109.5	9.5	258	9	US-09-813-453A-2	Sequence 2, Appli
19	108	9.4	219	9	US-09-813-453A-57	Sequence 57, Appl
20	105.5	9.2	257	9	US-09-813-453A-53	Sequence 53, Appl
21	102.5	8.9	250	9	US-09-813-453A-3	Sequence 3, Appli
22	102	8.8	212	9	US-09-813-453A-59	Sequence 59, Appl
23	101	8.8	256	9	US-09-813-453A-55	Sequence 55, Appl
24	100.5	8.7	255	9	US-09-813-453A-7	Sequence 7, Appli
25	98.5	8.5	262	9	US-09-813-453A-45	Sequence 45, Appl
26	91	7.9	254	9	US-09-813-453A-47	Sequence 47, Appl
27	89.5	7.8	272	9	US-09-712-363-276	Sequence 276, App
28	89.5	7.8	272	9	US-09-813-453A-5	Sequence 5, Appli
29	88.5	7.7	262	9	US-09-813-453A-11	Sequence 11, Appl
30	87.5	7.6	258	9	US-09-813-453A-49	Sequence 49, Appl
31	87	7.5	947	10	US-09-801-574-4	Sequence 4, Appli
32	84	7.3	764	10	US-09-815-242-11556	Sequence 11556, A
33	82.5	7.2	265	9	US-09-813-453A-4	Sequence 4, Appli
34	82	7.1	688	10	US-09-815-242-11416	Sequence 11416, A
35	81.5	7.1	273	9	US-09-813-453A-10	Sequence 10, Appl
36	80	6.9	260	9	US-09-813-453A-51	Sequence 51, Appl
37	79	6.9	310	10	US-09-815-242-10283	Sequence 10283, A
38	79	6.9	375	9	US-09-981-353-113	Sequence 113, App
39	79	6.9	2799	9	US-10-151-736-4	Sequence 4, Appli
40	78	6.8	770	9	US-10-153-668-7	Sequence 7, Appli
41	77.5	6.7	258	9	US-09-813-453A-6	Sequence 6, Appli
42	77	6.7	692	9	US-10-101-464A-897	Sequence 897, App
43	76	6.6	434	9	US-10-153-668-454	Sequence 454, App
44	76	6.6	439	9	US-09-374-046A-44	Sequence 44, Appl
45	76	6.6	653	9	US-10-174-590-402	Sequence 402, App

ALIGNMENTS

RESULT 1
US-09-895-913A-74
; Sequence 74, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456Alel Helicobacter Polypeptides in
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-74

Query Match 100.0%; Score 1153; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 1.8e-107;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
```

```
Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 2
US-09-813-453A-14
; Sequence 14, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-813-453A-14

```
  Query Match             100.0%;  Score 1153;  DB 9;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-107;
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 3
US-09-813-453A-67
; Sequence 67, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 67
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-813-453A-67

```
  Query Match             100.0%;  Score 1153;  DB 9;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-107;
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              ||||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 4
US-09-813-453A-21
; Sequence 21, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
US-09-813-453A-21

```
  Query Match             34.0%;  Score 391.5;  DB 9;  Length 209;
  Best Local Similarity   39.2%;  Pred. No. 2.3e-31;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::|||||:  :|  : : |:   :        :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
               :|  :| | ||||||  ||  : :|||||||||||||:|    ||||  ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
```

RESULT 5
US-09-813-453A-9
; Sequence 9, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers

```
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Thermotoga maritima
US-09-813-453A-9

  Query Match             14.3%;  Score 165;  DB 9;  Length 246;
  Best Local Similarity   23.2%;  Pred. No. 1.3e-08;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:           |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 ----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                       :: |:||   |  :|      |   | || :           :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                :|       ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

Qy        144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
              |  |   |: | | : :  |:|         :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
              | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 6
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

  Query Match             13.8%;  Score 159.5;  DB 9;  Length 248;
  Best Local Similarity   28.6%;  Pred. No. 4.6e-08;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
              :||  | | : |      :|       : | :||  :   ||  ||:|| :| :| ::

Db         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                  :|:| |:|:| ||:  :| |||| | ||:             :|  |  : |
Db        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
                || :|          | :||  | :| :         :|| : |  : :|:| ||||:
Db        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy        191 YLSAFLPHSVCKERLVFDGMEIA 213
               :   |  :     ||| |: :|
Db        222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 7
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             12.8%;  Score 148;  DB 9;  Length 249;
  Best Local Similarity   29.3%;  Pred. No. 6.6e-07;
  Matches   58;  Conservative   34;  Mismatches   78;  Indels   28;  Gaps    7;

Qy         34 SAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACL 93
              ||:|  | :    : | ||          |   |  :||  :   |:  ||:|| :| :
Db         58 SAEETAKLVCALADTFSISP---------VCAAPAPELAGVINGYDDFARLGLDRWLAFV 108

Qy         94 AVNNGV-----VVDAGSAITIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILE----QP 143
                : |      |:| |:||| | ::  |||||| | ||:     :  :   : :     :
Db        109 GAYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEA 168

Qy        144 FKALDSLEVLPKSTRDAVNYG--MVLSVIACIQ------HLAKNQKIYLCGGDAKYLSAF 195
               |||  | |  ::| :||  |  ::|   |  |      :   :  |:: ||||  ::
Db        169 EKALVRL-VPGRATAEAVERGCSLMLRGFAMTQIEIAREYWGDDFAIFVTGGDAVLVADV 227

Qy        196 LPHSVCKERLVFDGMEIA 213
               || :      ||| |: :|
Db        228 LPGARIVPDLVFVGLALA 245


RESULT 8
US-09-813-453A-63
; Sequence 63, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
```

```
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
;   LENGTH: 241
;   TYPE: PRT
;   ORGANISM: Thiobacillus ferrooxidans
US-09-813-453A-63

  Query Match             12.5%;  Score 144.5;  DB 9;  Length 241;
  Best Local Similarity   24.4%;  Pred. No. 1.4e-06;
  Matches   60;  Conservative   40;  Mismatches   97;  Indels   49;  Gaps   10;

Qy         13 LVLCDIGNTRIHFAQNYQ--LFSSAK--------EDLKRLGIQKEIFYISVNEENEKALL 62
              ::   :||||   |  :    | |          | |:: |:     ::|       ||
Db          1 MIFIAVGNTRTLLAHTHDGVHFDSVSVATSLPPTEILQQPGLT----WLSAPNREPVALG 56

Qy         63 NCYPNA--------------KNIAGFFH--LETDY---VGLGIDRQMACLAV------NN 97
                | |               :   |||    :  ||  .    || ||:   ||        :
Db         57 GVVPAALAAWREALATAEVREPDPGFFRRAVPHDYHPPESLGFDRRCCLLAAAMDYPGQD 116

Qy         98 GVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDS-LEVLPKS 156
               :|:| |:||||||:  |   || ||||:|  :    :  :|  |   |::  |:|
Db        117 SIVIDMGTAITIDLLAGGHFRGGRILPGIAMSLRGLHEGTALL--PEVVLNAPAEMLGND 174

Qy        157 TRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDG 209
              | :|:  |::          |   :  :   :| : ||||:      :  |: :  |:  |
Db        175 TSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRG 234

Qy        210 MEIALK 215
                : ::
Db        235 FYLWIR 240


RESULT 9
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             12.5%;  Score 144.5;  DB 9;  Length 249;
  Best Local Similarity   29.8%;  Pred. No. 1.5e-06;
  Matches   61;  Conservative   26;  Mismatches   63;  Indels   55;  Gaps   10;

Qy         51 ISVNEENEKALLNC--------------YPNAKNIAGFFHLETDYVGLGIDRQMAC---- 92
              :||  | |   | |                :|  : :||  :   ||  ||:|| :|
Db         54 VSVRSEEETDAL-CALIAQAFAVQAKVAHP-VREMAGVRNGYDDYQRLGMDRWLAALGAF 111

Qy         93 -LAVNNGVVVDAGSAITIDLIK-EGKHLGGCILPGL----------AQYIHAYKKSAKIL 140
               ||    :|:| |:|   | :  :|:|||| | ||:           : |     ||:
Db        112 HLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRIRYDDASAE-- 169

Qy        141 EQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ----------KIYLCGGD 188
                :|| ||    :|| :||  | ||     :|  |  |             ::| |||
Db        170 ----RALSSLSP-GRSTVEAVERGCVL----MLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        189 AKYLSAFLPHSVCKERLVFDGMEIA 213
                |  : | ||  :     ||| |: :|
Db        221 APLVRAALPQARVVPDLVFVGLAMA 245


RESULT 10
US-09-813-453A-13
; Sequence 13, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Synechocystis sp.
US-09-813-453A-13

  Query Match             12.3%;  Score 141.5;  DB 9;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 3.1e-06;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy         18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
              |||:|:|:|     : |    |:                ||    :    | |: |
Db         25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy         59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                        |:| ||:                  |||| :| |          :|||
Db         80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy        104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
              |:|:||    : | | || |||||  :       ::   |   :| |  ||      |
Db        124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy        159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
               |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||        : |
Db        182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

Qy        206 VFDGM 210
                :| ||
Db        242 IFLGM 246


RESULT 11
US-09-813-453A-12
; Sequence 12, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
```

```
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Aquifex aeolicus
US-09-813-453A-12

  Query Match             12.0%;  Score 138.5;  DB 9;  Length 229;
  Best Local Similarity   26.5%;  Pred. No. 5.2e-06;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy         17 DIGNTRIHFA------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
              |:||: :  |              : :        || ||       ||| :   : :
Db          7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy         64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                |  | :   | :: ||     || ||   :        | ||: ||:|: |||: |
Db         60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy        115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
              ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
Db        120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172

Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                : | :   : : | : |: : ||: || | |    :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219


RESULT 12
US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
;  SEQ ID NO 41
;   LENGTH: 244
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

  Query Match             11.2%;  Score 129.5;  DB 9;  Length 244;
  Best Local Similarity   26.0%;  Pred. No. 4.5e-05;
  Matches   58;  Conservative   40;  Mismatches   82;  Indels   43;  Gaps   12;

Qy         12 NLVLCDIGNT--RIHFAQNYQLFS------SAKEDLKRL----GIQKEIF-YISVNEENE 58
              ||:: | ||:  :: | :|  : |       | : |  |       | |:  : : :|
Db          4 NLIV-DQGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEA 62

Qy         59 KALLNCYPNAKNIAGF-----FHLETDYVGLGIDRQMACLAVNN------GVVVDAGSAI 107
              :|::     |  : |        |: |   || ||  | :  ::       :|:|||:||
Db         63 EAIVRSCAAASLMMGTETPVPLRLQYDRRTLGADRLAAVVGAHSLYPNTELLVIDAGTAI 122

Qy        108 TIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK------STRDA 160
              | : :  || :||| | |||    |   |:  :        :|   : ||      || :|
Db        123 TYERVSAEGIYLGGNISPGL----HLRFKALHLFTGRLPLIDPSGISPKIAEYGSSTEEA 178

Qy        161 VNYGMVLSVIACIQ------HLAKNQK-IYLCGGDAKYLSAFL 196
              :  |:: :   |        |  : :  : | |||| ||:  :
Db        179 ITAGVIHGLAGEIDRYIDDLHAKEGRSAVILTGGDANYLARII 221


RESULT 13
US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-22

  Query Match             10.1%;  Score 116.5;  DB 9;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0029;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:|: |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 14
US-09-813-453A-43
; Sequence 43, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-43
```

```
  Query Match             10.1%;  Score 116.5;  DB 9;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0029;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                 : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

          190 KYLSAFLPHSVCKERLV 206
                ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 15
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39

  Query Match             10.1%;  Score 116;  DB 9;  Length 460;
  Best Local Similarity   25.5%;  Pred. No. 0.0023;
  Matches   62;  Conservative   37;  Mismatches   88;  Indels   56;  Gaps   12;

Qy          9 DLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KEIFYISVNE 55
              | :  :| : ||:|: :|  :|     : |:   ||  || :         |   :|
Db        204 DSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCG 263

Qy         56 ENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-----AVNNG 98
              |::||         :   |::    |    : | |      | ||    |     : |
Db        264 ESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGSRRFSRNAC 319

Qy         99 VVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI----LEQPFKALDSLEVL 153
              |||  |:|:|:| |  :| :||| |:||     |  |:|  :    | :|
Db        320 VVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRP---AGKRYPF 372

Qy        154 PKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDAKYLSAFLPHSVCKE 203
              | :| :||  ||: :|   | :           |   : : || |  ::  || :   |
Db        373 PTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAE 432

Qy        204 RLV 206
              |
Db        433 NTV 435
```

```
Search completed: June 24, 2003, 22:29:10
Job time : 13.3931 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 19.6161 Seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453A-67
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	16	O25533	O25533 helicobacte
2	1115	96.7	223	16	Q9ZKY6	Q9zky6 helicobacte
3	391.5	34.0	209	16	Q9PIA9	Q9pia9 campylobact
4	165	14.3	246	16	Q9WZY5	Q9wzy5 thermotoga
5	159.5	13.8	248	16	Q9HWC1	Q9hwc1 pseudomonas
6	158	13.7	276	16	Q8YQD7	Q8yqd7 anabaena sp
7	141.5	12.3	257	16	P74045	P74045 synechocyst
8	139	12.1	256	16	Q8RFE4	Q8rfe4 fusobacteri
9	138.5	12.0	229	16	O67753	O67753 aquifex aeo
10	126.5	11.0	259	16	Q8XHL5	Q8xhl5 clostridium
11	117.5	10.2	273	16	Q97EB4	Q97eb4 clostridium
12	116.5	10.1	592	16	Q9JXF1	Q9jxf1 neisseria m
13	116.5	10.1	592	16	Q9JWI7	Q9jwi7 neisseria m
14	116	10.1	224	16	Q98Q93	Q98q93 mycoplasma
15	115.5	10.0	242	16	Q9PCI4	Q9pci4 xylella fas
16	102	8.8	212	2	O32514	O32514 desulfovibr
17	100.5	8.7	295	16	Q8Y2M4	Q8y2m4 ralstonia s
18	94.5	8.2	255	16	Q8R7M2	Q8r7m2 thermoanaer
19	94.5	8.2	338	5	O76462	O76462 drosophila
20	94	8.2	407	16	O31557	O31557 bacillus su
21	93.5	8.1	1164	10	Q9XIK6	Q9xik6 arabidopsis
22	91.5	7.9	440	17	Q8ZZ98	Q8zz98 pyrobaculum
23	91.5	7.9	683	11	Q9Z2P1	Q9z2p1 mus musculu
24	91	7.9	254	16	Q9KGH5	Q9kgh5 bacillus ha
25	90.5	7.8	274	16	Q9CD56	Q9cd56 mycobacteri
26	90.5	7.8	482	16	Q9X0V6	Q9x0v6 thermotoga
27	90	7.8	313	16	Q8XKL3	Q8xkl3 clostridium
28	89.5	7.8	272	16	O06282	O06282 mycobacteri
29	89	7.7	243	12	Q98237	Q98237 molluscum c
30	89	7.7	594	5	Q9VVK8	Q9vvk8 drosophila
31	89	7.7	1483	10	Q8S1U8	Q8s1u8 oryza sativ
32	88.5	7.7	258	2	Q9F985	Q9f985 bacillus st
33	88.5	7.7	262	16	O51477	O51477 borrelia bu
34	88.5	7.7	493	12	Q9PYQ7	Q9pyq7 xestia c-ni
35	88	7.6	274	16	Q8RI50	Q8ri50 fusobacteri
36	87.5	7.6	383	2	Q9FD71	Q9fd71 enterococcu
37	87.5	7.6	483	8	Q35316	Q35316 oryza sativ
38	87	7.5	947	11	Q99MW9	Q99mw9 mus musculu
39	87	7.5	1051	17	Q8TMS3	Q8tms3 methanosarc
40	86.5	7.5	393	16	Q8YRY5	Q8yry5 anabaena sp
41	86	7.5	474	16	Q9PQV7	Q9pqv7 ureaplasma
42	86	7.5	702	16	Q8UD86	Q8ud86 agrobacteri
43	85.5	7.4	240	16	Q8ZKR0	Q8zkr0 salmonella
44	85.5	7.4	240	16	Q8Z2X1	Q8z2x1 salmonella
45	85.5	7.4	401	16	Q9KFC0	Q9kfc0 bacillus ha

ALIGNMENTS

RESULT 1
O25533
ID O25533 PRELIMINARY; PRT; 223 AA.
AC O25533;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein HP0862.
GN HP0862.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL; AE000596; AAD07916.1; -.
DR TIGR; HP0862; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 223 AA; 24655 MW; C1D13E8BEC1EB9AF CRC64;

Query Match 100.0%; Score 1153; DB 16; Length 223;
Best Local Similarity 100.0%; Pred. No. 2.2e-94;

```
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223


RESULT 2
Q9ZKY6
ID   Q9ZKY6      PRELIMINARY;      PRT;   223 AA.
AC   Q9ZKY6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PUTATIVE.
GN   JHP0796.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001509; AAD06372.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   223 AA;  24754 MW;  B31377DA1AEF6E0E CRC64;

  Query Match             96.7%;  Score 1115;  DB 16;  Length 223;
  Best Local Similarity   94.2%;  Pred. No. 5e-91;
  Matches  210;  Conservative   11;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||| |||:|:||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFKDLKDLILCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||:|||||||||||||| |||:|||||||||||:||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYIGLGIDRQMACLAVVNGVIVDAGSAITIDLVKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              |||||||||:||||||||||||||||||||||:|||||||:|:|:|||||||:|
Db        121 CILPGLAQYVHAYKKSAKILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCIQHLAKDQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223


RESULT 3
Q9PIA9
ID   Q9PIA9      PRELIMINARY;      PRT;   209 AA.
AC   Q9PIA9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Cj0394c.
GN   CJ0394C.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC   Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
DR   EMBL; AL139075; CAB74230.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   209 AA;  23935 MW;  C49313AA9E3D2D19 CRC64;

  Query Match             34.0%;  Score 391.5;  DB 16;  Length 209;
  Best Local Similarity   39.2%;  Pred. No. 6.5e-27;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::||||||:  :|  : : |:   :        :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
              :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203


RESULT 4
Q9WZY5
ID   Q9WZY5      PRELIMINARY;      PRT;   246 AA.
AC   Q9WZY5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TM0883.
GN   TM0883.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
```

```
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001754; AAD35964.1; -.
DR   TIGR; TM0883; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   246 AA;  27154 MW;  9E0309AD462CF266 CRC64;

  Query Match           14.3%;  Score 165;  DB 16;  Length 246;
  Best Local Similarity  23.2%;  Pred. No. 9.7e-07;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:          |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 ----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                      :: |:||   |  :|      |   | || :           :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|       ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

Qy        144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
               |  |   |: | | : :  |:|         :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
               | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 5
Q9HWC1
ID   Q9HWC1      PRELIMINARY;      PRT;   248 AA.
AC   Q9HWC1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA4279.
GN   PA4279.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07667.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001230; Prenyl_site.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   248 AA;  26758 MW;  609D37480899DA3B CRC64;

  Query Match           13.8%;  Score 159.5;  DB 16;  Length 248;
  Best Local Similarity  28.6%;  Pred. No. 3e-06;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
              :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::
Db         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                    :|:| |:|:| ||:  :| |||| | ||:             :|  |  : |
Db        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
               || :|          | :||  | :| :          :|| : |  : :|:| ||||:
Db        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy        191 YLSAFLPHSVCKERLVFDGMEIA 213
               :  |  :   |||  |: :|
Db        222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 6
Q8YQD7
ID   Q8YQD7      PRELIMINARY;      PRT;   276 AA.
AC   Q8YQD7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Alr3896.
GN   ALR3896.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003594; BAB75595.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   276 AA;  30272 MW;  2AD04CA693E56F25 CRC64;

  Query Match           13.7%;  Score 158;  DB 16;  Length 276;
  Best Local Similarity  25.1%;  Pred. No. 4.7e-06;
  Matches   70;  Conservative   51;  Mismatches   94;  Indels   64;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLC-DIGNTRIHFA-----------------QNYQLFSSAKEDLKR 41
              | ::||   ::|  |  :|||:|:|:|                 |   : :  |:
Db          1 MISQQSAKHIENQWLALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEV 60

Qy         42 LGIQKEIF--------------YISVNEENEKALLNCYPNAKNIA-GFFHLETDYVGLG 85
                 :||||               :|:     :  |   | | : |      |   |  ||
Db         61 GSEEKEIFFTFFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPTLG 120

Qy         86 IDRQMACLAVNNG-----VVVDAGSAITIDLIKEGKHL-GGCILPGLA-QYIHAYKKSAK 138
              ||| :|           :|:|||:|:|      ||:| || ||||:  |:    ::: :
Db        121 IDRALALWGAGMSWGFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQ 180

Qy        139 ILEQPFKALDSLEVLPK---STRDAVNYGMVLSVIACIQ-------HLAKNQKIYLCGGD 188
              : :   :|: ||   |:   :| :|:  |:: ::|| ::         |  : |: : |||
Db        181 LPQVEMEAIKSLP--PRFALNTTEAIQSGVIYTLIAGMRDFTEEWLSLFPDGKVAIKGGD 238

Qy        189 ----AKYLSAFLP----HSVCKERLVFDGMEIALKKAGI 219
                  || |  |      : :  |:| ||:  :  ||:
Db        239 RILLLNYLQALYPDLAARLIVEPNLIFWGMQTIV--AGV 275


RESULT 7
P74045
```

```
ID   P74045      PRELIMINARY;      PRT;   257 AA.
AC   P74045;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein slr0812.
GN   SLR0812.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RT   DNA Res. 3:109-136(1996).
DR   EMBL; D90911; BAA18120.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   257 AA;  27702 MW;  4291E51EE91FB4A2 CRC64;

  Query Match             12.3%;  Score 141.5;  DB 16;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 0.00012;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy         18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
              |||:|:|:|     : |    |:             ||    :    | |: |
Db         25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy         59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                        |:| ||:                 |||| :| |          :|||
Db         80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy        104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
              |:|:||    : | | || |||||   :       ::   |   :| |  ||      |
Db        124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy        159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
               |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||       : |
Db        182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

          206 VFDGM 210
              :| ||
Db        242 IFLGM 246


RESULT 8
Q8RFE4
ID   Q8RFE4      PRELIMINARY;      PRT;   256 AA.
AC   Q8RFE4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bvg accessory factor.
GN   FN0761.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010586; AAL94957.1; -.
KW   Complete proteome.
SQ   SEQUENCE   256 AA;  28601 MW;  CDF681127F08914B CRC64;

  Query Match             12.1%;  Score 139;  DB 16;  Length 256;
  Best Local Similarity   25.9%;  Pred. No. 0.00021;
  Matches   59;  Conservative   31;  Mismatches   86;  Indels   52;  Gaps    8;

Qy         17 DIGNTRI---HFAQNYQLFSSAKEDLKRLGIQKEIF----------YISVNEENEKALLN 63
              ||||| |    :  | :| |: :        : | |          ||: : :   : :
Db          6 DIGNTHIVTGIYDNNGELISTFRIATNDKMTEDEYFSYFNNITKYNEISIKKVDAILISS 65

Qy         64 CYPN---------------------AKNIAGFFHLETDYVGLGIDR------QMACLAV 95
                ||                     | :   |    :| | | ||       |
Db         66 VVPNIIITFQFFARKYFKVEATIVDLEKKLPFTFAKGINYTGFGADRIIDITEAMQKYPD 125

Qy         96 NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQPFKALDSLEVLP 154
              | |: | |:| | |::|:| ::|| |||:   |:| |  :||:     |    |  ||
Db        126 KNLVIFDFGTATTYDVLKKGVYIGGGILPGIDMSINALYGNTAKLPRVKFTTPSS--VLG 183

Qy        155 KSTRDAVNYGMVLSVIACIQHLAK------NQKIYL--CGGDAKYLSA 194
              |   :   :       |:|: |      |::|::   ||  | |||
Db        184 TDTMKQIQAAIFFGYAGQIKHIIKKINEELNEEIFVLATGGLGKILSA 231


RESULT 9
O67753
ID   O67753      PRELIMINARY;      PRT;   229 AA.
AC   O67753;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein AQ_1924.
GN   AQ_1924.
OS   Aquifex aeolicus.
OC   Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX   NCBI_TaxID=63363;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VF5;
RX   MEDLINE=98196666; PubMed=9537320;
RA   Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA   Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA   Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT   "The complete genome of the hyperthermophilic bacterium Aquifex
RT   aeolicus.";
RL   Nature 392:353-358(1998).
DR   EMBL; AE000763; AAC07720.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   229 AA;  26068 MW;  64C347A285FABE9A CRC64;

  Query Match             12.0%;  Score 138.5;  DB 16;  Length 229;
  Best Local Similarity   26.5%;  Pred. No. 0.0002;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy         17 DIGNTRIHFA----------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
              |:||: :  |            : :        || ||       ||| :   : :
Db          7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy         64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
              |  | :   | :: ||     || ||     :       | ||: ||:|: |||: |
Db         60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy        115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
              ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
```

```
Db        120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172

Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                  : | :   : : | : |: : ||: || | |     :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219


RESULT 10
Q8XHL5
ID   Q8XHL5      PRELIMINARY;      PRT;   259 AA.
AC   Q8XHL5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CPE2468.
GN   CPE2468.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003194; BAB82174.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28819 MW;  0D5FEA3B7A145E10 CRC64;

  Query Match             11.0%;  Score 126.5;  DB 16;  Length 259;
  Best Local Similarity   22.8%;  Pred. No. 0.0027;
  Matches   63;  Conservative   43;  Mismatches   77;  Indels   93;  Gaps   13;

Qy         13 LVLCDIGNTR----IHFAQNY----------------------QLFSSAK---EDLKRLG 43
              ::| |:|||     ||  : |                      |||: ||   ||:: :
Db          2 ILLIDVGNTNIVLGIHDNEKYIASWRISTDSKKTSDEYSIQVMQLFNQAKLNPEDVEGII 61

Qy         44 IQKEIFYISVNEENEKALLNC----------------YPNAKNIAGFFHLETDYVGLGI 86
              |   :  |  : ||     |                | | | :             |
Db         62 ISSVVPNIMHSLENMVRKCFCKEPIVVGPGIKTGINIKYDNPKEV------------GA 108

Qy         87 DRQMACLAV-----NNGVVVDAGSAITIDLIKE-GKHLGGCILPGLAQYIHA-YKKSAKI 139
              || :  :|        :::| |:| |    | | | :||| | ||:      | ::::||:
Db        109 DRIVNAVAAFEKHKKPMIIIDFGTATTFCAITEKGDYLGGNICPGIQISADALFERAAKL 168

Qy        140 ----LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQK-----------IY 183
                  ||:|        |: |:|  ::  |::   |  :::: |  |           :
Db        169 PRIELEKP------KSVICKNTVTSMQAGIIYGYIGKVEYIVKRMKKEMMDLGEKEPFVL 222

Qy        184 LCGGDAKYLSAFLPHSVCKE---RLVFDGMEIALKK 216
                || ||  :  :    |  |   :|  :|::|  :|
Db        223 ATGGLAKLV--YSETDVIDEVDRKLTLEGLKILYEK 256


RESULT 11
Q97EB4
ID   Q97EB4      PRELIMINARY;      PRT;   273 AA.
AC   Q97EB4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Predicted transcriptional regulator, homolog of Bvg accessory
DE   factor.
GN   CAC3200.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007815; AAK81136.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30331 MW;  DE7B4D5923E72842 CRC64;

  Query Match             10.2%;  Score 117.5;  DB 16;  Length 273;
  Best Local Similarity   21.4%;  Pred. No. 0.018;
  Matches   57;  Conservative   48;  Mismatches   94;  Indels   67;  Gaps   11;

Qy         10 LKNLVLCDIGNTRI--------HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKAL 61
              || ::: |:||| |             ::| :          |||  :  :   :: :  |
Db         10 LKVILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQ--VMNLFQQDKLDPTL 67

Qy         62 LN-----------CYPNAKNIAGFFHLETDYVGLGI---------------DRQMACLA 94
              :             |     |  :| :    || ||                || :  :|
Db         68 VEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVA 127

Qy         95 VN-----NGVVVDAGSAITIDLIKE-GKHLGGCILPGLAQYIHA-YKKSAKI----LEQP 143
              :     : :::| |:| |   ::| | :||| | ||:     | ::|:||:    | :|
Db        128 AHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKP 187

Qy        144 FKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQK-----------IYLCGGDAKY 191
              |     : |:|  ::  |:|   |  :::: :  |                :   || ||
Db        188 AYA------ICKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKL 241

Qy        192 LSAFLPH-SVCKERLVFDGMEIALKK 216
              :|    :  |    |  :|: |  :|
Db        242 ISEEAKNVDVINPFLTLEGLRIIYEK 267


RESULT 12
Q9JXF1
ID   Q9JXF1      PRELIMINARY;      PRT;   592 AA.
AC   Q9JXF1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein/Bvg accessory factor.
GN   NMB2075.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
```

```
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002557; AAF42394.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; NMB2075; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  64701 MW;  9E27421DA2B41DE3 CRC64;

  Query Match             10.1%;  Score 116.5;  DB 16;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.059;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
              ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 13
Q9JWI7
ID   Q9JWI7      PRELIMINARY;      PRT;   592 AA.
AC   Q9JWI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
     01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
     01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein (EC 6.3.4.15).
GN   BIRA OR NMA0357.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162752; CAB83659.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   592 AA;  64468 MW;  7ED2E0CD3B31C630 CRC64;

  Query Match             10.1%;  Score 116.5;  DB 16;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.059;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||         :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
              ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 14
Q98Q93
ID   Q98Q93      PRELIMINARY;      PRT;   224 AA.
AC   Q98Q93;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   Hypothetical protein MYPU_4730.
GN   MYPU_4730.
OS   Mycoplasma pulmonis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC   Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2107;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=UAB CTIP;
RX   MEDLINE=21267165; PubMed=11353084;
RA   Chambaud I., Heilig R., Ferris S., Barbe V., Samson D., Galisson F.,
RA   Moszer I., Dybvig K., Wroblewski H., Viari A., Rocha E.P.C.,
RA   Blanchard A.;
RT   "The complete genome sequence of the murine respiratory pathogen
RT   Mycoplasma pulmonis.";
RL   Nucleic Acids Res. 29:2145-2153(2001).
DR   EMBL; AL445564; CAC13646.1; -.
DR   MypuList; MYPU_4730; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   224 AA;  25363 MW;  240600CED6AB2B4B CRC64;

  Query Match             10.1%;  Score 116;  DB 16;  Length 224;
  Best Local Similarity   22.3%;  Pred. No. 0.019;
  Matches   51;  Conservative   30;  Mismatches   94;  Indels   54;  Gaps    7;

Qy         14 VLCDIGNTRIHFA-----------------------QNYQLFSSAKEDLKRLGIQKEI 48
              :  |:||:  ||                        ::: ||   | |:     |||
Db          3 IFIDLGNSFAKFALKNENKAHFLFRLKTQSVVDPSFEIKSFNLFEFNKLDV------KEI 56

Qy         49 FYISVNEENEKALL-----NCYPNAKNIAGFFHLE----------TDYVGLGI--DRQM 90
```

```
Db         57 LICSVRNAKENQILEFKLKSIFKNAK-IDFFIHKSQSLVKFCQKELTSEIGLDIVANAYY 115

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150

Db        116 VLHKSKNAIFISLGTAIVITQIKNSSIEGISIYPGIYQSFKNFFNVVAKIESNFNIIKIP 175

Qy        151 EVLPKSTRDAVN----YGMVLSVIACIQHLAKNQKIYLCGGDAKYLSAF 195

Db        176 PILGKTTLESISSSLVRGSVFLLKGYIDEIDKTSDIFITGGTITFSKIF 224


RESULT 15
Q9PCI4
ID   Q9PCI4      PRELIMINARY;      PRT;   242 AA.
AC   Q9PCI4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Xf1795.
GN   XF1795.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004001; AAF84603.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   242 AA;  25658 MW;  EDC5DE7C0449C1E9 CRC64;

  Query Match             10.0%;  Score 115.5;  DB 16;  Length 242;
  Best Local Similarity   31.5%;  Pred. No. 0.023;
  Matches   45;  Conservative   22;  Mismatches   55;  Indels   21;  Gaps    6;

Qy         85 GIDRQMACL---AVNNGVVVDAGSAITIDLI-KEGKHLGGCILPG---LAQYIHAYKKSA 137

Db         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152

Qy        138 KILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACI--------QHLAKNQKIYLCGGDA 189

Db        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211

Qy        190 KYLSAFLPHSVCKERLVFDGMEI 212

Db        212 ASLLPWLGDVVHRPTLVLDGLAI 234


Search completed: June 24, 2003, 22:16:25
Job time : 20.6161 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 9.04773 Seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453A-67
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	89.5	7.8	1582	2	US-08-404-531B-9	Sequence 9, Appli
2	89.5	7.8	1582	3	US-08-476-900A-9	Sequence 9, Appli
3	89.5	7.8	1582	3	US-08-488-546A-9	Sequence 9, Appli
4	89.5	7.8	1582	4	US-08-726-320-5	Sequence 5, Appli
5	89.5	7.8	1582	4	US-09-208-716-5	Sequence 5, Appli
6	86.5	7.5	1581	2	US-08-404-531B-6	Sequence 6, Appli
7	86.5	7.5	1581	3	US-08-476-900A-6	Sequence 6, Appli
8	86.5	7.5	1581	3	US-08-488-546A-6	Sequence 6, Appli
9	86.5	7.5	1581	4	US-08-726-320-4	Sequence 4, Appli
10	86.5	7.5	1581	4	US-09-208-716-4	Sequence 4, Appli
11	81	7.0	1580	4	US-08-726-320-1	Sequence 1, Appli
12	81	7.0	1580	4	US-09-208-716-1	Sequence 1, Appli
13	81	7.0	1581	4	US-08-726-320-3	Sequence 3, Appli
14	81	7.0	1581	4	US-09-208-716-3	Sequence 3, Appli
15	79.5	6.9	424	6	5169835-6	Patent No. 5169835
16	78	6.8	382	4	US-09-134-001C-5449	Sequence 5449, Ap
17	77.5	6.7	1447	4	US-09-376-330-17	Sequence 17, Appl
18	77	6.7	601	4	US-09-134-001C-3791	Sequence 3791, Ap
19	76	6.6	374	2	US-07-857-224B-82	Sequence 82, Appl
20	76	6.6	374	2	US-07-857-224B-83	Sequence 83, Appl
21	75.5	6.5	375	4	US-09-347-878-56	Sequence 56, Appl
22	74.5	6.5	248	4	US-09-081-689-4	Sequence 4, Appli
23	74.5	6.5	248	4	US-08-858-207A-340	Sequence 340, App
24	74.5	6.5	442	4	US-09-081-689-2	Sequence 2, Appli
25	74.5	6.5	442	4	US-09-305-984-14	Sequence 14, Appl
26	74.5	6.5	442	4	US-09-073-541A-14	Sequence 14, Appl
27	74	6.4	513	4	US-09-134-001C-4490	Sequence 4490, Ap
28	73.5	6.4	904	6	5244792-4	Patent No. 5244792
29	72.5	6.3	903	1	US-08-220-151-8	Sequence 8, Appli
30	72.5	6.3	903	1	US-08-413-118-8	Sequence 8, Appli
31	72.5	6.3	903	3	US-08-473-446-8	Sequence 8, Appli
32	72.5	6.3	1012	2	US-08-680-326-34	Sequence 34, Appl
33	72.5	6.3	1462	3	US-07-792-600-31	Sequence 31, Appl
34	72.5	6.3	1462	3	US-09-157-021-31	Sequence 31, Appl
35	72.5	6.3	1462	3	US-09-156-842-31	Sequence 31, Appl
36	72	6.2	248	6	5169835-15	Patent No. 5169835
37	71.5	6.2	307	4	US-08-804-166-4	Sequence 4, Appli
38	71.5	6.2	307	4	US-08-910-991-4	Sequence 4, Appli
39	70.5	6.1	511	2	US-09-019-201A-2	Sequence 2, Appli
40	70.5	6.1	579	1	US-08-126-564A-31	Sequence 31, Appl
41	70.5	6.1	579	5	PCT-US94-09143-31	Sequence 31, Appl
42	70	6.1	355	1	US-08-008-688A-2	Sequence 2, Appli
43	70	6.1	613	2	US-08-915-207-2	Sequence 2, Appli
44	70	6.1	613	2	US-08-915-207-4	Sequence 4, Appli
45	70	6.1	613	4	US-09-238-555-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-08-404-531B-9
; Sequence 9, Application US/08404531B
; Patent No. 5863724
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT:  Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 5863724
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5863724ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/404,531B
;      FILING DATE:  15-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1582 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-404-531B-9

  Query Match             7.8%;  Score 89.5;  DB 2;  Length 1582;
  Best Local Similarity   22.7%;  Pred. No. 0.83;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223
```

```
Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |          :| |   |  :  ::|    ::  ||| :|| |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 2
US-08-476-900A-9
; Sequence 9, Application US/08476900A
; Patent No. 6031150
  GENERAL INFORMATION:
;   APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
;   TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6031150
;   TITLE OF INVENTION: and Method of Detecting Persistent Hyperinsulinemic Hypoglycem
;   TITLE OF INVENTION: Infancy
;   NUMBER OF SEQUENCES: 49
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
;     ADDRESSEE: No. 6031150ris
;     STREET: One Liberty Place 46th. Floor
;     CITY: Philadelphia
;     STATE: PA
;     COUNTRY: USA
;     ZIP: 19103
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/476,900A
;     FILING DATE: 07-JUN-1995
;     CLASSIFICATION: 800
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Beardell, Lori Y.
;     REGISTRATION NUMBER: 34,293
;     REFERENCE/DOCKET NUMBER: BYLR-0027
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 215-568-3100
;     TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1582 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-476-900A-9

  Query Match             7.8%;  Score 89.5;  DB 3;  Length 1582;
  Best Local Similarity   22.7%;  Pred. No. 0.83;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
               |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |          :| |   |  :  ::|    ::  ||| :|| |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 3
US-08-488-546A-9
; Sequence 9, Application US/08488546A
; Patent No. 6054313
;  GENERAL INFORMATION:
;    APPLICANT: Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT: Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION: Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6054313
;    NUMBER OF SEQUENCES: 49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE: No. 6054313ris
;      STREET: One Liberty Place 46th. Floor
;      CITY: Philadelphia
;      STATE: PA
;      COUNTRY: USA
;      ZIP: 19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE: Floppy disk
;      COMPUTER: IBM PC compatible
;      OPERATING SYSTEM: PC-DOS/MS-DOS
;      SOFTWARE: PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER: US/08/488,546A
;      FILING DATE: 07-JUNE-1995
;      CLASSIFICATION: 800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 08/404,531
;      FILING DATE: 15-MARCH-1995
;      CLASSIFICATION: 800
;    ATTORNEY/AGENT INFORMATION:
;      NAME: Beardell, Lori Y.
;      REGISTRATION NUMBER: 34,293
;      REFERENCE/DOCKET NUMBER: BYLR-0026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE: 215-568-3100
;      TELEFAX: 215-568-3439
;  INFORMATION FOR SEQ ID NO: 9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH: 1582 amino acids
;      TYPE: amino acid
;      TOPOLOGY: linear
;    MOLECULE TYPE: protein
US-08-488-546A-9

  Query Match             7.8%;  Score 89.5;  DB 3;  Length 1582;
  Best Local Similarity   22.7%;  Pred. No. 0.83;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
               |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337
```

```
Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |         :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
               : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411
```

RESULT 4
US-08-726-320-5
; Sequence 5, Application US/08726320
; Patent No. 6171815
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Coleman, Roger
; TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/726,320
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: AF-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 784874
US-08-726-320-5

Query Match 7.8%; Score 89.5; DB 4; Length 1582;
Best Local Similarity 22.7%; Pred. No. 0.83;
Matches 58; Conservative 44; Mismatches 81; Indels 73; Gaps 15;

```
Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337
```

```
Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |         :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
               : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411
```

RESULT 5
US-09-208-716-5
; Sequence 5, Application US/09208716
; Patent No. 6235880
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Coleman, Roger
; TITLE OF INVENTION: NOVEL HUMAN SULFONYLUREA RECEPTOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,716
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/726,320
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: AF-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 784874
US-09-208-716-5

Query Match 7.8%; Score 89.5; DB 4; Length 1582;
Best Local Similarity 22.7%; Pred. No. 0.83;
Matches 58; Conservative 44; Mismatches 81; Indels 73; Gaps 15;

```
Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
```

```
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 6
US-08-404-531B-6
; Sequence 6, Application US/08404531B
; Patent No. 5863724
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT:  Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 5863724
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 5863724ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/404,531B
;      FILING DATE:  15-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0003
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-404-531B-6

  Query Match             7.5%;  Score 86.5;  DB 2;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 7
US-08-476-900A-6
; Sequence 6, Application US/08476900A
; Patent No. 6031150
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor
; Patent No. 6031150
;    TITLE OF INVENTION:  and Method of Detecting Persistent Hyperinsulinemic Hypogly
;    TITLE OF INVENTION:  Infancy
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 6031150ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/476,900A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0027
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-476-900A-6

  Query Match             7.5%;  Score 86.5;  DB 3;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 8
US-08-488-546A-6
; Sequence 6, Application US/08488546A
```

```
; Patent No. 6054313
;  GENERAL INFORMATION:
;    APPLICANT:  Joseph Bryan, Lydia Aguilar Bryan, Daniel Nelson, Pamela
;    APPLICANT:  Thomas, Gilbert Cote, and Robert Gagel
;    TITLE OF INVENTION:  Sequence Encoding Mammalian Sulfonylurea Receptor          a
; Patent No. 6054313
;    NUMBER OF SEQUENCES:  49
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock Washburn Kurtz Mackiewicz &
;      ADDRESSEE:  No. 6054313ris
;      STREET:  One Liberty Place 46th. Floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/488,546A
;      FILING DATE:  07-JUNE-1995
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/404,531
;      FILING DATE:  15-MARCH-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Beardell, Lori Y.
;      REGISTRATION NUMBER:  34,293
;      REFERENCE/DOCKET NUMBER:  BYLR-0026
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  215-568-3100
;      TELEFAX:  215-568-3439
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-488-546A-6

  Query Match             7.5%;  Score 86.5;  DB 3;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

 y         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||:::|    ::  :|   |  : |
Db       1173 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1223

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |          :  :|:|  :    |  :::|         |:|
Db       1224 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1281

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    | |  : ||   |:     : :: | :  ::||
Db       1282 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1337

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |          :| |   |  :  ::|    ::  ||| :|| |  |·    |: |
Db       1338 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1395

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1396 VDMFEGRIIIDGIDIA 1411


RESULT 9
US-08-726-320-4
; Sequence 4, Application US/08726320
; Patent No. 6171815
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,320
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  Genbank
;      CLONE:  1311534
US-08-726-320-4

  Query Match             7.5%;  Score 86.5;  DB 4;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |  |  |  ||:::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |          :  :|:|  :    |  :::|         |:|
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1280

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    | |  : ||   |:     : :: | :  ::||
Db       1281 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1336

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |          :| |   |  :  ::|    ::  ||| :|| |  |    |: |
Db       1337 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1394

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1395 VDMFEGRIIIDGIDIA 1410


RESULT 10
US-09-208-716-4
; Sequence 4, Application US/09208716
; Patent No. 6235880
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
```

```
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,716
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,320
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  Genbank
;      CLONE:  1311534
US-09-208-716-4

  Query Match             7.5%;  Score 86.5;  DB 4;  Length 1581;
  Best Local Similarity   21.5%;  Pred. No. 1.8;
  Matches   55;  Conservative   45;  Mismatches   83;  Indels   73;  Gaps   14;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | |   |  |  ||:::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |         :  :|:|  :    |  :::|         |:|
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1280

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:         ::|    |  |  : ||   |:     : :: | :  ::||
Db       1281 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKGIHTLLKTE---AESYEGLLAPSLIPK 1336

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
              :  |         :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1337 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRM 1394

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1395 VDMFEGRIIIDGIDIA 1410


RESULT 11
US-08-726-320-1
; Sequence 1, Application US/08726320
; Patent No. 6171815
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,320
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1580 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-08-726-320-1

  Query Match             7.0%;  Score 81;  DB 4;  Length 1580;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||:::|    ::  :|   |  : |                | ::
Db       1174 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1231

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |         :  :|:|  :    |  :::|         |:|   |  :|  |:
Db       1232 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1288

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                  ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1289 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1345

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1346 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1403

Qy        208 DGMEIA 213
                ||::||
Db       1404 DGIDIA 1409


RESULT 12
US-09-208-716-1
; Sequence 1, Application US/09208716
; Patent No. 6235880
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
```

```
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,716
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,320
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1580 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:
;      CLONE:  Consensus
US-09-208-716-1

  Query Match             7.0%;  Score 81;  DB 4;  Length 1580;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                 | ::
Db       1174 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1231

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |         :  :|:|  :    |  :::|         |:|   |  :|  |:
Db       1232 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1288

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                 ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1289 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1345

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1346 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1403

Qy        208 DGMEIA 213
               ||::||
Db       1404 DGIDIA 1409


RESULT 13
US-08-726-320-3
; Sequence 3, Application US/08726320
; Patent No. 6171815
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/726,320
;      FILING DATE:  Filed Herewith
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  AF-0001 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1369844
US-08-726-320-3

  Query Match             7.0%;  Score 81;  DB 4;  Length 1581;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------------YPNA 68
              :| | |  |  |  ||::|    ::  :|   |  : |                 | ::
Db       1175 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1232

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |         :  :|:|  :    |  :::|         |:|   |  :|  |:
Db       1233 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1289

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                 ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1290 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1346

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1347 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1404

Qy        208 DGMEIA 213
               ||::||
Db       1405 DGIDIA 1410


RESULT 14
US-09-208-716-3
; Sequence 3, Application US/09208716
; Patent No. 6235880
;  GENERAL INFORMATION:
;    APPLICANT:  Au-Young, Janice
;    APPLICANT:  Bandman, Olga
;    APPLICANT:  Coleman, Roger
;    TITLE OF INVENTION:  NOVEL HUMAN SULFONYLUREA RECEPTOR
;    NUMBER OF SEQUENCES:  5
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Incyte Pharmaceuticals, Inc.
;      STREET:  3174 Porter Drive
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.
;      ZIP:  94304
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/208,716
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/726,320
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J.
;      REGISTRATION NUMBER:  36,749
       REFERENCE/DOCKET NUMBER:  AF-0001 US
     TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-855-0555
;      TELEFAX:  415-845-4166
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1581 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    IMMEDIATE SOURCE:
;      LIBRARY:  GenBank
;      CLONE:  1369844
US-09-208-716-3

  Query Match             7.0%;  Score 81;  DB 4;  Length 1581;
  Best Local Similarity   21.1%;  Pred. No. 7.5;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC---------------YPNA 68
              .  :| | |  |  |  ||::|    ::  :|   |  : |                 | ::
Db       1175 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1232

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
                |||  |         :  :|:|  :    |  :::|        |:|   |  :|  |:
Db       1233 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1289

           113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRD------ 159
                    ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1290 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1346

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  :  ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1347 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1404

Qy        208 DGMEIA 213
               ||::||
Db       1405 DGIDIA 1410


RESULT 15
5169835-6
;Patent No. 5169835
;    APPLICANT: WAI-YEE, CHAN
;    TITLE OF INVENTION: PREGNANCY SPECIFIC PROTEINS APPLICATIONS
;    NUMBER OF SEQUENCES: 48
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/390,409
;      FILING DATE: 07-AUG-1989
;SEQ ID NO:6:
;      LENGTH: 424
5169835-6

  Query Match             6.9%;  Score 79.5;  DB 6;  Length 424;
  Best Local Similarity   22.3%;  Pred. No. 1.5;
  Matches   40;  Conservative   23;  Mismatches   61;  Indels   55;  Gaps    7;

Qy         56 ENEKALLNCYPNAKNIAGF----------FHLETDYVGLGIDRQM---------------- 90
              | :  ||  :   :|: |:          :|  | ||   :| |:
Db         47 EGKDVLLLVHNLPQNLTGYIWYKGQMTDLYHYITSYV---VDGQIIYGPAYSGRETVYSN 103

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
              | | : |    ||||  |: :|| |   ||          :  |       | |  :: |
Db        104 ASLLIQNVTQEDAGS-YTLHIIKRGDGTGG---------VTGYFTVTLYSETPKPSISSS 153

Qy        151 EVLPKSTRDAVNY----------------GMVLSVIACIQHLAKNQKIYLCGGDAKYLS 193
               : |:   :||                 |  | :  :|    |: :|| . |  ||::
Db        154 NLNPREVMEAVRLICDPETPDASYLWLLNGQNLPMTHRLQLSKTNRTLYLF-GVTKYIA 211


Search completed: June 24, 2003, 22:10:53
Job time : 10.0477 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 4.48585 Seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453A-67
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	109.5	9.5	233	1	YACB_BACSU	P37564 bacillus su
2	89.5	7.8	232	1	PYRK_ARCFU	O29722 archaeoglob
3	89.5	7.8	1581	1	ACC8_CRICR	Q09427 cricetus cr
4	88.5	7.7	1182	1	DP3A_RICPR	O05974 rickettsia
5	88.5	7.7	1581	1	ACC8_RAT	Q09429 rattus norv
6	87.5	7.6	344	1	HOLA_HAEIN	P43747 haemophilus
7	85.5	7.4	426	1	GLPC_HAEIN	P43801 haemophilus
8	84	7.3	764	1	SYFB_HELPJ	Q9zkf8 helicobacte
9	83	7.2	303	1	HIS1_HAEIN	P43853 haemophilus
10	82.5	7.2	349	1	GPDA_DROAE	Q27556 drosophila
11	82.5	7.2	352	1	GPDA_DROVI	P07735 drosophila
12	82.5	7.2	424	1	PSGA_HUMAN	Q15235 homo sapien
13	81.5	7.1	293	1	NPL_HAEIN	P44539 haemophilus
14	81.5	7.1	314	1	ARCC_CLOPE	Q46171 clostridium
15	81	7.0	1580	1	ACC8_HUMAN	Q09428 homo sapien
16	81	7.0	2452	1	RPB1_PLAFD	P14248 plasmodium
17	80.5	7.0	353	1	GPDA_DROPS	Q27928 drosophila
18	80	6.9	253	1	YTBQ_BACSU	P53560 bacillus su
19	80	6.9	327	1	MTH1_HAEHA	P05102 haemophilus
20	80	6.9	449	1	MURF_RICCN	Q92h60 rickettsia
21	80	6.9	558	1	VB04_VACCC	P21001 vaccinia vi
22	80	6.9	558	1	VB04_VACCV	P24769 vaccinia vi
23	80	6.9	633	1	YR45_CAEEL	Q09562 caenorhabdi
24	79.5	6.9	370	1	ALR2_SCHPO	Q9p5n3 schizosacch
25	79	6.9	310	1	ARCL_ECOLI	Q46807 escherichia
26	79	6.9	374	1	ADHA_HUMAN	P07327 homo sapien
27	79	6.9	773	1	GYRB_HELPJ	Q9zlx3 helicobacte
28	79	6.9	1065	1	CARB_CLOAB	Q97ft3 clostridium
29	79	6.9	2799	1	EDD_HUMAN	O95071 homo sapien
30	78.5	6.8	131	1	YIL2_STAAU	P41370 staphylococ
31	78.5	6.8	349	1	GPDA_DROEZ	Q27567 drosophila
32	78.5	6.8	355	1	ALR_RICCN	Q92jd9 rickettsia
33	78.5	6.8	359	1	GPDA_DROKA	O97463 drosophila
34	78.5	6.8	500	1	CPN1_CAVPO	Q64408 cavia porce
35	78	6.8	425	1	CLPX_RICCN	Q92gq4 rickettsia
36	78	6.8	717	1	NTP2_MSEPV	Q9yw06 melanoplus
37	78	6.8	920	1	EDD_RAT	Q62671 rattus norv
38	77.5	6.7	1448	1	UGGG_SCHPO	Q09140 schizosacch
39	77	6.7	389	1	PSD6_ORYSA	Q8w425 oryza sativ
40	77	6.7	917	1	VGLB_HSVB2	P12641 bovine herp
41	76.5	6.6	326	1	Y338_CHLMU	Q9pkx2 chlamydia m
42	76.5	6.6	335	1	G3P_BORBU	P46795 borrelia bu
43	76.5	6.6	484	1	MURE_CLOPE	Q8xj99 clostridium
44	76.5	6.6	508	1	YI57_THEMA	Q9x2h2 thermotoga
45	76.5	6.6	538	1	THSB_METTH	O26885 methanobact

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             9.5%;  Score 109.5;  DB 1;  Length 233;
  Best Local Similarity   26.7%;  Pred. No. 0.021;
  Matches   35;  Conservative   24;  Mismatches   47;  Indels   25;  Gaps    5;

Qy         74 FFHLETDYVG----------------LGIDRQMACLAV-----NNGVVVDAGSAITIDLI 112
              :||:|   ||                :| || :  :|      |  :||| |:| |   |
Db         80 YFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYI 139

Qy        113 KEGK-HLGGCILPGLAQYIHA-YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVI 170
              | | ::|| | ||:     | | ::||:        |:  :: |:|  |:  |::   :
Db        140 DENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDN--IIGKNTVSAMQSGILFGYV 197

Qy        171 ACIQHLAKNQK 181
              ::  : |  |
Db        198 GQVEGIVKRMK 208


RESULT 2
PYRK_ARCFU
ID   PYRK_ARCFU     STANDARD;      PRT;   232 AA.
AC   O29722;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable dihydroorotate dehydrogenase electron transfer subunit.
GN   PYRK OR AF0528.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
     SEQUENCE FROM N.A.
     STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- FUNCTION: Is responsible for channelling the electrons from the
CC       oxidation of dihydroorotate from the FMN redox center in the pyrD
CC       subunit to the ultimate electron acceptor NAD(+) (By similarity).
CC   -!- COFACTOR: Binds a 2Fe-2S cluster and FAD (By similarity).
CC   -!- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC   -!- SUBUNIT: Heterotetramer of 2 pyrK and 2 pyrD subunits (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE PYRK FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001068; AAB90709.1; -.
DR   TIGR; AF0528; -.
DR   InterPro; IPR000564; 2Fe2S_ferredoxin.
DR   InterPro; IPR001834; Cyt_B5_reductase.
DR   InterPro; IPR001433; Oxred_FAD/NAD(P).
DR   Pfam; PF00175; NAD_binding; 1.
DR   Pfam; PF00970; FAD_binding_6; 1.
DR   PROSITE; PS00197; 2FE2S_FERREDOXIN; 1.
KW   Pyrimidine biosynthesis; Transport; Electron transport; Metal-binding;
KW   Iron; Iron-sulfur; Flavoprotein; FAD; Complete proteome.
FT   METAL       202    202       IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
FT   METAL       207    207       IRON-SULFUR 1 (2FE-2S) (BY SIMILARITY).
FT   METAL       210    210       IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
FT   METAL       219    219       IRON-SULFUR 2 (2FE-2S) (BY SIMILARITY).
SQ   SEQUENCE   232 AA;  25639 MW;  907C6A5D3098784B CRC64;

  Query Match             7.8%;  Score 89.5;  DB 1;  Length 232;
  Best Local Similarity   24.7%;  Pred. No. 1.1;
  Matches   54;  Conservative   36;  Mismatches   74;  Indels   55;  Gaps   14;

Qy         39 LKRLGIQKEIFYISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNG 98
              ::||  :    | |::       |  ||      | | :   ||  | :     |: :|
Db          9 IERLTDEVATLYFSIS-------LRSYP------GQFVMV--YVP-GCEEIPLSLSSSNS 52

Qy         99 VVVDAGSAITIDLI--KEGKHL------GGCILP-----------GLA--QYIHAY---- 133
              | | |     |  ||  |||:::      |   |           |:|  :|:: |
Db         53 VTVKAVGETTASLINAKEGQYVGVRGAFGSAFTPSKRALIVAGGIGIAPMKYLYEYLMKC 112

Qy        134 --KKSAKILEQPFKA---LDSLEVLPKSTRDAVNYGMVLSVIAC--IQHLAKNQKIYLCG 186
                | |    |:  |    ||  : :  :| |  ::|:  :|:    :: | : :|||:||
Db        113 GSKVSVVYGERTAKNLFWLDKFDRITVTTEDG-SFGLKGTVLDALKLEKLDEYEKIYVCG 171

Qy        187 GDAKYLSAFLPHSVCKERLVFDGMEIALK---KAGILEC 222
                :    : :      |||   |  | :|:   : ||  |
Db        172 SEGMLRATY---DFLKERDALDKAEFSLERYMRCGIGVC 207


RESULT 3
ACC8_CRICR
ID   ACC8_CRICR     STANDARD;      PRT;  1581 AA.
AC   Q09427;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Sulfonylurea receptor 1.
GN   ABCC8 OR SUR.
OS   Cricetus cricetus (Black-bellied hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetus.
OX   NCBI_TaxID=10034;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=95232532; PubMed=7716547;
RA   Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV,
RA   Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J.,
RA   Nelson D.A.;
RT   "Cloning of the beta cell high-affinity sulfonylurea receptor: a
RT   regulator of insulin secretion.";
RL   Science 268:423-426(1995).
RN   [2]
RP   CARBOHYDRATE-LINKAGE SITE ASN-9.
```

```
RX   MEDLINE=97098100; PubMed=8942641;
RA   Nelson D.A., Bryan J., Wechsler S., Clement J.P. IV, Aguilar-Bryan L.;
RT   "The high-affinity sulfonylurea receptor: distribution, glycosylation,
RT   purification, and immunoprecipitation of two forms from endocrine and
RT   neuroendocrine cell lines.";
RL   Biochemistry 35:14793-14799(1996).
RN   [3]
RP   TOPOLOGY, AND CARBOHYDRATE-LINKAGE SITE ASN-9.
RX   MEDLINE=99436117; PubMed=10506167;
RA   Raab-Graham K.F., Cirilo L.J., Boettcher A.A., Radeke C.M.,
RA   Vandenberg C.A.;
RT   "Membrane topology of the amino-terminal region of the sulfonylurea
RT   receptor.";
RL   J. Biol. Chem. 274:29122-29129(1999).
CC   -!- FUNCTION: PUTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC       POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC       AND INSULIN RELEASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40623; AAA99201.1; -.
DR   HSSP; P13569; 1NBD.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABCtranprtrTM.
DR   InterPro; IPR000388; Surea_receptor.
DR   Pfam; PF00005; ABC_tran; 2.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   PRINTS; PR01092; SULFNYLUREAR.
DR   ProDom; PD000006; ABC_transportr; 2.
DR   SMART; SM00382; AAA; 2.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   Receptor; Transmembrane; ATP-binding; Transport; Glycoprotein; Repeat;
KW   Phosphorylation.
FT   INIT_MET      0      0
FT   DOMAIN        1     33       EXTRACELLULAR.
FT   TRANSMEM     34     54       1.
FT   DOMAIN       55     74       CYTOPLASMIC.
FT   TRANSMEM     75     95       2.
FT   DOMAIN       96    100       EXTRACELLULAR.
FT   TRANSMEM    101    121       3.
FT   DOMAIN      122    133       CYTOPLASMIC.
FT   TRANSMEM    134    153       4.
FT   DOMAIN      154    166       EXTRACELLULAR.
FT   TRANSMEM    167    193       5.
FT   DOMAIN      194    310       CYTOPLASMIC.
FT   TRANSMEM    311    330       6 (BY SIMILARITY).
FT   DOMAIN      331    354       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    355    375       7 (BY SIMILARITY).
FT   DOMAIN      376    433       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    434    454       8 (BY SIMILARITY).
FT   DOMAIN      455    457       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    458    478       9 (BY SIMILARITY).
FT   DOMAIN      479    540       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    541    561       10 (BY SIMILARITY).
FT   DOMAIN      562    583       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    584    604       11 (BY SIMILARITY).
FT   DOMAIN      605   1004       CYTOPLASMIC.
FT   TRANSMEM   1005   1025       12.
FT   DOMAIN     1026   1072       EXTRACELLULAR.
FT   TRANSMEM   1073   1093       13 (BY SIMILARITY).
FT   DOMAIN     1094   1137       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1138   1158       14 (BY SIMILARITY).
FT   DOMAIN     1159   1159       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1160   1180       15 (BY SIMILARITY).
FT   DOMAIN     1181   1251       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1252   1272       16 (BY SIMILARITY).
FT   DOMAIN     1273   1276       EXTRACELLULAR.
FT   TRANSMEM   1277   1297       17.
FT   DOMAIN     1298   1581       CYTOPLASMIC.
FT   NP_BIND     712    719       ATP (POTENTIAL).
FT   NP_BIND    1378   1385       ATP (POTENTIAL).
FT   CARBOHYD      9      9       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .).
SQ   SEQUENCE   1581 AA;  177015 MW;  0EFCCFB3A454E0B1 CRC64;

  Query Match             7.8%;  Score 89.5;  DB 1;  Length 1581;
  Best Local Similarity   22.7%;  Pred. No. 9.6;
  Matches   58;  Conservative   44;  Mismatches   81;  Indels   73;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|        :| | |  |  |  ||::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLVSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF------HLET--DYVGLGIDRQMACLAVNN--------GVVVD 102
                  | :: |||  |        ||   :|:|  :    |  :::|         |:|
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVCMEYIGACVVLIAAATSISNSLHRELSAGLV-- 1280

Qy        103 AGSAITIDLIKEG------KHLGGC-ILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK 155
               |  :|  |:          ::|    |  |  : |||  |:     : :: | :  ::||
Db       1281 -GLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHALLKTE---AESYEGLLAPSLIPK 1336

Qy        156 STRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLP 197
               :  |          :| |   |  :  ::|    ::  ||| :||  |  |    |: |
Db       1337 NWPDQGKIQIQNLSVRYDSSLKPV--LKHVNTLISPGQKIGICGRTGSGKSSFSLAFFRM 1394

Qy        198 HSVCKERLVFDGMEIA 213
                : : |:: ||::||
Db       1395 VDMFEGRIIIDGIDIA 1410


RESULT 4
DP3A_RICPR
ID   DP3A_RICPR     STANDARD;      PRT;  1182 AA.
AC   O05974;
DT   15-DEC-1998 (Rel. 37, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA polymerase III alpha subunit (EC 2.7.7.7).
GN   DNAE OR RP778.
OS   Rickettsia prowazekii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=782;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=99039499; PubMed=9823893;
RA   Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA   Sicheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA   Eriksson A.-S., Winkler H.H., Kurland C.G.;
RT   "The genome sequence of Rickettsia prowazekii and the origin of
RT   mitochondria.";
RL   Nature 396:133-140(1998).
RN   [2]
RP   SEQUENCE OF 1-780 FROM N.A.
RC   STRAIN=Madrid E;
RX   MEDLINE=97419517; PubMed=9274032;
RA   Andersson J.O., Andersson S.G.E.;
RT   "Genomic rearrangements during evolution of the obligate
RT   intracellular parasite Rickettsia prowazekii as inferred from an
RT   analysis of 52015 bp nucleotide sequence.";
RL   Microbiology 143:2783-2795(1997).
RN   [3]
RP   SEQUENCE OF 822-1182 FROM N.A.
RC   STRAIN=Madrid E, and B;
RX   MEDLINE=99416441; PubMed=10486973;
RA   Andersson J.O., Andersson S.G.E.;
```

```
RT   "Genome degradation is an ongoing process in Rickettsia.";
RL   Mol. Biol. Evol. 16:1178-1191(1999).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC       THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC       CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC       DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC       THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC       AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
     modified and this statement is not removed. Usage by and for commercial
     entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ235273; CAA15204.1; -.
DR   EMBL; Y11785; CAA72479.1; -.
DR   EMBL; AJ238755; CAB56087.1; -.
DR   EMBL; AJ238756; CAB56091.1; -.
DR   InterPro; IPR004013; PHP_C.
DR   InterPro; IPR003141; PHP_N.
DR   InterPro; IPR004805; PolC_alpha.
DR   InterPro; IPR004365; tRNA_anti.
DR   Pfam; PF01336; tRNA_anti; 1.
DR   Pfam; PF02231; PHP_N; 1.
DR   Pfam; PF02811; PHP_C; 1.
DR   SMART; SM00481; POLIIIAc; 1.
DR   TIGRFAMs; TIGR00594; polc; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   Complete proteome.
FT   CONFLICT    376    376       I -> V (IN REF. 2).
SQ   SEQUENCE   1182 AA;  133462 MW;  0FE9B56969C2E52A CRC64;

  Query Match             7.7%;  Score 88.5;  DB 1;  Length 1182;
  Best Local Similarity   21.3%;  Pred. No. 8.4;
  Matches   59;  Conservative   35;  Mismatches  104;  Indels   79;  Gaps    9;

Qy          4 RQSFTDLKNLVLCDIG-NTRIHFA----------------QNYQLFSSAKED-LKRLGI 44
              |: |:|| | :   |    | :||                : | :   | |  | ||
Db        243 RKLFSDLPNAIQNTINLRERCYFAAHPNPPMLPNFSTQDISETYLIRKYATEGLLARLVT 302

Qy         45 QKEIFYISVNEENEKALLNCYPNAK---------NIAGFFHLETDYVGLGIDRQMACLAV 95
              : :   ||: |  ||     :             | ||:| : :|::     :   : |
Db        303 KFKAENISL-EHQEKLKTEYFTRLNYELDIICNMNFAGYFLIVSDFIKWS---KKQGILV 358

Qy         96 NNGVVVDAGSAIT-------IDLIKEG-------------------------------- 115
                |   ||| :         :| || |
Db        359 GPGRGSGAGSVVAWSLLITDLDPIKFGLLFERFLNPERISMPDFDIDFCQERREEVINYV 418

Qy        116 ------KHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL-EVLPKSTRDAVNYGMVLS 168
                   :|  |  |  |     |  |::|  |:|  | | |::| |  : |:    :
Db        419 RSKYGPNRVGQIITFGKMQAKAVIKDVARVLSLPYKLADYLTELVPFSAINPVSLEQAIR 478

Qy        169 VIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERL 205
               :  : : ||   :|   |||: :   |  |:  | |
Db        479 EVPELANAAKGNGLYNLEGDAELIKLVLDTSLILEGL 515


RESULT 5
ACC8_RAT
ID   ACC8_RAT       STANDARD;      PRT;  1581 AA.
AC   Q09429; O54989; P70532; Q9EQT0;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sulfonylurea receptor 1.
GN   ABCC8 OR SUR1 OR SUR.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM C).
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=95232532; PubMed=7716547;
RA   Aguilar-Bryan L., Nichols C.G., Wechsler S.W., Clement J.P. IV,
RA   Boyd A.E. III, Gonzalez G., Herrera-Sosa H., Nguy K., Bryan J.,
RA   Nelson D.A.;
RT   "Cloning of the beta cell high-affinity sulfonylurea receptor: a
RT   regulator of insulin secretion.";
RL   Science 268:423-426(1995).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RC   TISSUE=Pancreatic islets;
RA   Faure C., Partiseti M., Gouhier C., Graham D.;
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM B).
RC   TISSUE=Pancreatic islets;
RA   Blache P., Peyrollier K., Gros L., Bataille D.;
RT   "Identification of two forms of sulfonylurea receptors (SUR 1A and SUR
RT   1B) by means of polymerase chain reaction, subcloning, DNA sequencing
RT   and tissue distribution.";
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RC   TISSUE=Vascular smooth muscle;
RA   Cao K., Wang R.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   CARBOHYDRATE-LINKAGE SITE ASN-9.
RX   MEDLINE=97098100; PubMed=8942641;
RA   Nelson D.A., Bryan J., Wechsler S., Clement J.P. IV, Aguilar-Bryan L.;
RT   "The high-affinity sulfonylurea receptor: distribution, glycosylation,
RT   purification, and immunoprecipitation of two forms from endocrine and
RT   neuroendocrine cell lines.";
RL   Biochemistry 35:14793-14799(1996).
CC   -!- FUNCTION: PUTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC       POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC       AND INSULIN RELEASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 3 isoforms; A (shown here), B and C; may be
CC       produced by alternative splicing.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L40624; AAA99237.1; -.
DR   EMBL; X97279; CAA65934.1; -.
DR   EMBL; AF039595; AAB96684.1; -.
DR   EMBL; AB052294; BAB19011.1; -.
DR   HSSP; P13569; 1NBD.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001140; ABCtranprtrTM.
DR   InterPro; IPR000388; Surea_receptor.
DR   Pfam; PF00005; ABC_tran; 2.
DR   Pfam; PF00664; ABC_membrane; 2.
DR   PRINTS; PR01092; SULFNYLUREAR.
DR   ProDom; PD000006; ABC_transportr; 2.
```

```
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW   Receptor; Transmembrane; ATP-binding; Transport; Glycoprotein; Repeat;
KW   Phosphorylation; Alternative splicing.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     33       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM     34     54       1 (BY SIMILARITY).
FT   DOMAIN       55     74       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM     75     95       2 (BY SIMILARITY).
FT   DOMAIN       96    100       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    101    121       3 (BY SIMILARITY).
FT   DOMAIN      122    133       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    134    153       4 (BY SIMILARITY).
FT   DOMAIN      154    167       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    168    193       5 (BY SIMILARITY).
FT   DOMAIN      194    310       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    311    330       6 (BY SIMILARITY).
FT   DOMAIN      331    354       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    355    375       7 (BY SIMILARITY).
FT   DOMAIN      376    433       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    434    454       8 (BY SIMILARITY).
FT   DOMAIN      455    457       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    458    478       9 (BY SIMILARITY).
FT   DOMAIN      479    540       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM    541    561       10 (BY SIMILARITY).
FT   DOMAIN      562    583       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM    584    604       11 (BY SIMILARITY).
FT   DOMAIN      605   1004       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1005   1025       12 (BY SIMILARITY).
FT   DOMAIN     1026   1072       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1073   1093       13 (BY SIMILARITY).
FT   DOMAIN     1094   1137       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1138   1158       14 (BY SIMILARITY).
FT   DOMAIN     1159   1159       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1160   1180       15 (BY SIMILARITY).
FT   DOMAIN     1181   1251       CYTOPLASMIC (BY SIMILARITY).
FT   TRANSMEM   1252   1272       16 (BY SIMILARITY).
FT   DOMAIN     1273   1276       EXTRACELLULAR (BY SIMILARITY).
FT   TRANSMEM   1277   1297       17 (BY SIMILARITY).
FT   DOMAIN     1298   1581       CYTOPLASMIC (BY SIMILARITY).
FT   NP_BIND     712    719       ATP (POTENTIAL).
FT   NP_BIND    1378   1385       ATP (POTENTIAL).
FT   CARBOHYD      9      9       N-LINKED (GLCNAC. . .).
FT   CARBOHYD   1049   1049       N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT   VARSPLIC    740    740       MISSING (IN ISOFORM C).
FT   VARSPLIC   1251   1288       MISSING (IN ISOFORM B).
FT   CONFLICT    486    486       S -> T (IN REF. 1).
FT   CONFLICT    698    698       I -> T (IN REF. 4).
FT   CONFLICT    835    836       QR -> PG (IN REF. 1).
FT   CONFLICT   1313   1313       R -> G (IN REF. 1).
SQ   SEQUENCE   1581 AA;  177053 MW;  C554B1A48B2D3901 CRC64;

  Query Match             7.7%;  Score 88.5;  DB 1;  Length 1581;
  Best Local Similarity   21.2%;  Pred. No. 12;
  Matches   56;  Conservative   48;  Mismatches   71;  Indels   89;  Gaps   15;

Qy         14 VLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC--------- 64
              |:|       :| | | |  |  | ||::|    ::  :|   |  : |
Db       1172 VVC-------YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARF 1222

Qy         65 ------YPNAKNIAGFF--------HLETDYVGLGIDRQMACLAVNNGVVVDAGSAITID 110
                  | :: |||  |          :  :|:|       ||:     |:: | ::|:
Db       1223 QQKLLEYTDSNNIASLFLTAANRWLEVRMEYIG-------ACV-----VLIAAATSISNS 1270

Qy        111 LIKEGKHLGGCILPGL------AQYIH--------------AYKKSAKILE---QPFKAL 147
              | :|     | :  ||      : |::              | |:   :|:   : :: |
Db       1271 LHRELS--AGLVGLGLTYALMVSNYLNWMVRNLADMEIQLGAVKRIHTLLKTEAESYEGL 1328

Qy        148 DSLEVLPKSTRD---------AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK-- 190
               :  ::||:  |          :| |   | :  ::|    ::  ||| :||  |  |
Db       1329 LAPSLIPKNWPDQGKIQIQNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSS 1386

Qy        191 -YLSAFLPHSVCKERLVFDGMEIA 213
```

```
                |: |    : : |:: ||::||
Db       1387 FSLAFFRMVDMFEGRIIIDGIDIA 1410


RESULT 6
HOLA_HAEIN
ID   HOLA_HAEIN     STANDARD;      PRT;   344 AA.
AC   P43747;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   DNA polymerase III, delta subunit (EC 2.7.7.7).
GN   HOLA OR HI0923.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC       RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC       THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC       THE DELTA SUBUNIT SEEMS TO INTERACT WITH THE GAMMA SUBUNIT TO
CC       TRANSFER THE BETA SUBUNIT ON THE DNA (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC       + {DNA}(N).
CC   -!- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC       CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC       DIMERIZATION TO FORM THE POLIII' COMPLEX. POLIII' ASSOCIATES WITH
CC       THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC       AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32774; AAC22583.1; -.
DR   TIGR; HI0923; -.
DR   TIGRFAMs; TIGR01128; holA; 1.
KW   Transferase; DNA-directed DNA polymerase; DNA replication;
KW   Complete proteome.
SQ   SEQUENCE   344 AA;  39845 MW;  5048DF5A42A97B5F CRC64;

  Query Match             7.6%;  Score 87.5;  DB 1;  Length 344;
  Best Local Similarity   24.2%;  Pred. No. 2.5;
  Matches   29;  Conservative   24;  Mismatches   44;  Indels   23;  Gaps    5;

Qy         43 GIQKEIFYISVNE-ENEKALLNCY-PNAKNIAGFFHLETDYVGLGIDR---QMACLAVNN 97
              ||:|: ::|::|: |    |:|| |  :|:  :    |  :||  |    | | :  |
Db       119 GIEKQTWFITLNQYEPNTILINCQTPTVENLPRWVKNRTKAMGLDADNEAIQQLCYSYEN 178

Qy         98 GVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQ-----PFKALDSLEV 152
                ::    :   :||:     |              | :   ::||     ||: :|:| |
Db       179 NLLALKQALQLLDLLYPDHKLN------------YNRVISVVEQSSIFTPFQWIDALLV 225
```

```
RESULT 7
GLPC_HAEIN
ID   GLPC_HAEIN     STANDARD;      PRT;   426 AA.
AC   P43801;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Anaerobic glycerol-3-phosphate dehydrogenase subunit C (G-3-P
DE   dehydrogenase).
GN   GLPC OR HI0683.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- FUNCTION: ELECTRON TRANSFER PROTEIN; MAY ALSO FUNCTION AS THE
CC       MEMBRANE ANCHOR FOR THE GLPAB DIMER (BY SIMILARITY).
CC   -!- PATHWAY: FIRST STEP IN THE TRANSFER OF REDUCING EQUIVALENTS
CC       FROM G-3-P TO A SHORT ELECTRON TRANSFER CHAIN TERMINATING WITH
CC       FUMARATE AS THE ULTIMATE ELECTRON ACCEPTOR.
CC   -!- SUBUNIT: COMPOSED OF A CATALYTIC GLPA/B DIMER AND OF GLPC
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: LOOSELY BOUND TO THE CYTOPLASMIC MEMBRANE
CC       OFTEN OCCURRING IN VESICLES ASSOCIATED WITH FUMARATE REDUCTASE
CC       (BY SIMILARITY).
CC   -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC       'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32751; AAC22343.1; -.
DR   TIGR; HI0683; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR004017; DUF224.
DR   Pfam; PF00037; fer4; 1.
DR   Pfam; PF02754; DUF224; 2.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 2.
KW   Electron transport; Iron-sulfur; 4Fe-4S; Membrane; Complete proteome.
FT   METAL        32     32       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        35     35       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        38     38       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        42     42       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        79     79       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        82     82       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        85     85       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        89     89       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
SQ   SEQUENCE   426 AA;  48068 MW;  A6EA5F33C6C4FC3C CRC64;

  Query Match             7.4%;  Score 85.5;  DB 1;  Length 426;
  Best Local Similarity   30.4%;  Pred. No. 4.8;
  Matches   34;  Conservative   17;  Mismatches   40;  Indels   21;  Gaps    6;

Qy          56 ENEKA-----LLNCYPN-AKNIAGFFHLETDYVGLGIDRQMACLAVNNGV-VVDAGSAIT 108
               | ||       ::| :|| |:||| |    |||:|  :|         ||: |:   |: :
Db         226 EKEKCCGLPLMVNGFPNRARNIAQF---NTDYIGKMVD--------ENGIDVISEASSCS 274

Qy         109 IDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDA 160
               ::|  |  |: |     :  :||        | | ||   :| : |   | |
Db         275 LNLRDEYHHILGIDNAKVRPHIHMV---TPFLYQLFKEGKTLPLKPLKLRVA 323


RESULT 8
SYFB_HELPJ
ID   SYFB_HELPJ     STANDARD;      PRT;   764 AA.
AC   Q9ZKF8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Phenylalanyl-tRNA synthetase beta chain (EC 6.1.1.20) (Phenylalanine--
DE   tRNA ligase beta chain) (PheRS).
GN   PHET OR JHP0979.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + L-phenylalanine + tRNA(Phe) = AMP +
CC       diphosphate + L-phenylalanyl-tRNA(Phe).
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC       SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE PHENYLALANYL-TRNA SYNTHETASE BETA
CC       CHAIN FAMILY. SUBFAMILY 1.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001527; AAD06558.1; -.
DR   InterPro; IPR005146; B3_4.
DR   InterPro; IPR005147; B5.
DR   InterPro; IPR005121; Fdx-AntiCB.
DR   InterPro; IPR004532; PheT_bact.
DR   InterPro; IPR002547; tRNA_bind.
DR   Pfam; PF01588; tRNA_bind; 1.
DR   Pfam; PF03147; FDX-ACB; 1.
DR   Pfam; PF03483; B3_4; 1.
DR   Pfam; PF03484; B5; 1.
DR   TIGRFAMs; TIGR00472; pheT_bact; 1.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
SQ   SEQUENCE   764 AA;  85154 MW;  70BDAFD202FED425 CRC64;

  Query Match             7.3%;  Score 84;  DB 1;  Length 764;
  Best Local Similarity   21.5%;  Pred. No. 12;
  Matches   53;  Conservative   36;  Mismatches   85;  Indels   72;  Gaps   12;

Qy           2 PARQSFTDLKNLVLC----------DIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYI 51
               |       |:  ::|          ::|   |   :   :::| :|::::||     | :
Db         519 PITTELNTLRTSLVCGLLDASLRNKNLGFKSIALYEKGSVYNSKREEIQKLG-----FLV 573
```

```
Qy         52 SVNEENEKALLNCYPNAK--------------NIAGFFHLE--TDYVGLGIDRQMACLAV 95
              |  :: |      ||:||               | | | ||  |    :    | | :
Db        574 SGLQKKES-----YPHAKGKAWDFYSFAECVSRIIGDFSLEKLTTQTPINHPYQSAKIIQ 628

Qy         96 NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKK-----SAKILEQPFKALDSL 150
              || :       | :| :     | | : | :  ::       |  |::|   |
Db        629 NNEI---------IGVIAK-------IHPKVIQELDLFESYYAEIDASKLKRPAMLLKPF 672

Qy        151 EVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKY--LSAFLPHSVCKERLVFD 208
               : | | ||     :::        : |  |      ||:   ||  ||  : ||
Db        673 SIYPSSVRDLT---LIIDENTAFSKIKKALK------DAQIPNLSEILPLDIFKE----S 719

Qy        209 GMEIAL 214
              |  |||
Db        720 GNTIAL 725


RESULT 9
HIS1_HAEIN
ID   HIS1_HAEIN     STANDARD;      PRT;   303 AA.
AC   P43853;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   ATP phosphoribosyltransferase (EC 2.4.2.17).
GN   HISG OR HI0468.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: 1-(5-phospho-D-ribosyl)-ATP + diphosphate =
CC       ATP + 5-phospho-alpha-D-ribose 1-diphosphate.
CC   -!- PATHWAY: Histidine biosynthesis; first step. Very important in the
CC       regulation of histidine metabolism.
CC   -!- SUBUNIT: Homohexamer (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ATP PHOSPHORIBOSYLTRANSFERASE FAMILY.
CC       LONG SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32729; AAC22127.1; -.
DR   TIGR; HI0468; -.
DR   InterPro; IPR001348; HisG.
DR   Pfam; PF01634; HisG; 1.
DR   ProDom; PD003516; HisG; 1.
DR   TIGRFAMs; TIGR00070; hisG; 1.
DR   PROSITE; PS01316; ATP_P_PHORIBOSYLTR; 1.
KW   Histidine biosynthesis; Transferase; Glycosyltransferase;
KW   Complete proteome.

SQ   SEQUENCE   303 AA;  33821 MW;  08C14D1F6E98A31D CRC64;

  Query Match             7.2%;  Score 83;  DB 1;  Length 303;
  Best Local Similarity   23.7%;  Pred. No. 5.3;
  Matches   55;  Conservative   34;  Mismatches   79;  Indels   64;  Gaps   12;

Qy         11 KNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKN 70
              | |   | |: |:                   | : ::  | :| :   : :   ||
Db         98 KKLRQLDFGDCRL----------------SLAVDRDCHYENVKDLANRRIATSYP---- 136

Qy         71 IAGFFHLETDYVGLGIDRQMACLAVNNGVVV--DAGSAITI-DLIKEGKHLGGCILPGLA 127
                 ||   |:        :|| :|  | |   || |  | ||:  |  |      ||
Db        137 ----HLLKRYMNENGVSFKSCL-LNGSVEVAPSAGIAYAICDLVSSGATLEA---NGLK 187

Qy        128 QYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGG 187
              :   |:  | ::::     : ||    ||: |:    :|: |  :|  |::: | |
Db        188 EVDVIYRSKACLIQR----AEPLE----STKQAL-VDKLLTRIQGVQQAAESKYIMLHAP 238

Qy        188 DAKY--LSAFLP------------------HSVCKERLVFDGMEIALKKAG 218
                |   ::| ||                  | | :| | :: ||  ||:||
Db        239 KEKLEKITALLPGVENPTILPLASDTTRVAMHVVSQENLFWETME-QLKEAG 289


RESULT 10
GPDA_DROAE
ID   GPDA_DROAE     STANDARD;      PRT;   349 AA.
AC   Q27556;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic (EC 1.1.1.8)
DE   (GPD-C) (GPDH-C).
GN   GPDH.
OS   Drosophila americana (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=40366;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96025618; PubMed=8522168;
RA   Tominaga H., Narise S.;
RT   "Sequence evolution of the Gpdh gene in the Drosophila virilis
RT   species group.";
RL   Genetica 96:293-302(1995).
CC   -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC       phosphate + NADH.
CC   -!- PATHWAY: THIS ENZYME IS ONE OF THE TWO ENZYMES CONSTITUTING THE
CC       ALPHA-GLYCEROPHOSPHATE CYCLE, IN WHICH THIS ENZYME OPERATES IN
CC       CONJUNCTION WITH A MITOCHONDRIAL ALPHA-GLYCEROPHOSPHATE OXIDASE.
CC       THE ALPHA-GLYCEROPHOSPHATE CYCLE IS ESSENTIAL FOR THE PRODUCTION
CC       OF ENERGY FOR FLIGHT IN INSECTS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC       DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D50089; BAA20286.1; -.
DR   FlyBase; FBgn0015636; Dame\Gpdh.
DR   InterPro; IPR001652; NAD_Gly3P_dh.
DR   Pfam; PF01210; NAD_Gly3P_dh; 1.
DR   PRINTS; PR00077; GPDHDRGNASE.
DR   ProDom; PD001649; NAD_Gly3P_dh; 1.
DR   PROSITE; PS00957; NAD_G3PDH; 1.
KW   Oxidoreductase; NAD.
```

```
FT   INIT_MET      0      0       BY SIMILARITY.
SQ   SEQUENCE   349 AA;  38191 MW;  8D8C2F5C56E0BDB6 CRC64;

  Query Match             7.2%;  Score 82.5;  DB 1;  Length 349;
  Best Local Similarity   25.7%;  Pred. No. 6.8;
  Matches   43;  Conservative   27;  Mismatches   54;  Indels   43;  Gaps   10;

Qy         10 LKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAK 69
              |||:| |  |     |    :|  : |  : |||: : | :: |           || :|
Db        204 LKNIVACGAG-----FVDGLKLGDNTKAAVIRLGLMEMIRFVDV----------FYPGSK 248

Qy         70 NIAGFFH------LETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKE---GKHLGG 120
              :: ||       | |   | | :|:::   | :|  ::       || ||   |: | |
Db        249 -LSTFFESCGVADLITTCYG-GRNRRVSEAFVTSGKTIE-------DLEKEMLNGQKLQG 299

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSL------EVLPKSTRDAV 161
                |  |: :: |    | ||  |    ::      :: ||   | :
Db        300 ---PPTAEEVN-YMLKNKGLEDKFPLFTAIHKICTNQLKPKDLIDCI 342


RESULT 11
DA_DROVI
ID   GPDA_DROVI     STANDARD;      PRT;   352 AA.
AC   P07735; Q27634;
DT   01-AUG-1988 (Rel. 08, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glycerol-3-phosphate dehydrogenase [NAD+], cytoplasmic (EC 1.1.1.8)
DE   (GPD-C) (GPDH-C).
GN   GPDH.
OS   Drosophila virilis (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC   Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC   Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7244;
RN   [1]
RP   SEQUENCE.
RA   Arai K., Tominaga H., Yokote Y., Narise S.;
RT   "The complete amino-acid sequence of cytoplasmic glycerol-3-phosphate
RT   dehydrogenase from Drosophila virilis.";
RL   Biochim. Biophys. Acta 953:6-13(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Narise S., Tominaga H.;
RT   "Temperature-dependency differencies in GPDH allozmyes associated
RT   with a single base change in the coding region of the GPDH
RT   structural gene in Drosophila virilis.";
RL   Life Sci. Adv. (Genet.) 11:39-45(1992).
RN   [3]
     SEQUENCE FROM N.A.
     MEDLINE=92305069; PubMed=1610907;
RA   Tominaga H., Shiba T., Narise S.;
RT   "Structure of Drosophila virilis glycerol-3-phosphate dehydrogenase
RT   gene and a comparison with the Drosophila melanogaster gene.";
RL   Biochim. Biophys. Acta 1131:233-238(1992).
CC   -!- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC       phosphate + NADH.
CC   -!- PATHWAY: THIS ENZYME IS ONE OF THE TWO ENZYMES CONSTITUTING THE
CC       ALPHA-GLYCEROPHOSPHATE CYCLE, IN WHICH THIS ENZYME OPERATES IN
CC       CONJUNCTION WITH A MITOCHONDRIAL ALPHA-GLYCEROPHOSPHATE OXIDASE.
CC       THE ALPHA-GLYCEROPHOSPHATE CYCLE IS ESSENTIAL FOR THE PRODUCTION
CC       OF ENERGY FOR FLIGHT IN INSECTS.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC       DEHYDROGENASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X59076; CAA41800.1; -.
DR   EMBL; D10697; BAA01539.1; -.
DR   PIR; JS0023; JS0023.
DR   FlyBase; FBgn0013081; Dvir\Gpdh.
DR   InterPro; IPR001652; NAD_Gly3P_dh.
DR   Pfam; PF01210; NAD_Gly3P_dh; 1.
DR   PRINTS; PR00077; GPDHDRGNASE.
DR   ProDom; PD001649; NAD_Gly3P_dh; 1.
DR   PROSITE; PS00957; NAD_G3PDH; 1.
KW   Oxidoreductase; NAD.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       BLOCKED.
FT   CONFLICT     14     14       W -> N (IN REF. 1).
FT   CONFLICT     34     34       E -> K (IN REF. 2).
SQ   SEQUENCE   352 AA;  38520 MW;  FF7EB736962456F6 CRC64;

  Query Match             7.2%;  Score 82.5;  DB 1;  Length 352;
  Best Local Similarity   25.7%;  Pred. No. 6.9;
  Matches   43;  Conservative   27;  Mismatches   54;  Indels   43;  Gaps   10;

Qy         10 LKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAK 69
              |||:| |  |     |    :|  : |  : |||: : | :: |           || :|
Db        204 LKNIVACGAG-----FVDGLKLGDNTKAAVIRLGLMEMIRFVDV----------FYPGSK 248

Qy         70 NIAGFFH------LETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKE---GKHLGG 120
              :: ||       | |   | | :|:::   | :|  ::       || ||   |: | |
Db        249 -LSTFFESCGVADLITTCYG-GRNRRVSEAFVTSGKTIE-------DLEKEMLNGQKLQG 299

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSL------EVLPKSTRDAV 161
                |  |: :: |    | ||  |    ::      :: ||   | :
Db        300 ---PPTAEEVN-YMLKNKGLEDKFPLFTAIHKICTNQLKPKDLIDCI 342


RESULT 12
PSGA_HUMAN
ID   PSGA_HUMAN     STANDARD;      PRT;   424 AA.
AC   Q15235;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Pregnancy-specific beta-1-glycoprotein 10 precursor (PSBG-10).
GN   PSG10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=91104939; PubMed=2271648;
RA   Barnett T.R., Pickle W. II, Elting J.J.;
RT   "Characterization of two new members of the pregnancy-specific beta
RT   1-glycoprotein family from the myeloid cell line KG-1 and suggestion
RT   of two distinct classes of transcription unit.";
RL   Biochemistry 29:10213-10218(1990).
CC   -!- SUBCELLULAR LOCATION: Secreted (Potential).
CC   -!- DEVELOPMENTAL STAGE: PSBG ARE PRODUCED IN HIGH QUANTITY DURING
CC       PREGNANCY.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CEA
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; X17098; CAA34957.1; -.
DR   Genew; HGNC:9515; PSG10.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 4.
DR   SMART; SM00410; IG_like; 3.
DR   SMART; SM00408; IGc2; 1.
KW   Immunoglobulin domain; Glycoprotein; Signal; Repeat; Multigene family.
FT   SIGNAL        1     34       BY SIMILARITY.
FT   CHAIN        35    424       PREGNANCY-SPECIFIC BETA-1-GLYCOPROTEIN
FT                                10.
FT   DOMAIN       35    143       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      161    223       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      254    316       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      346    400       IG-LIKE C2-TYPE DOMAIN 3.
FT   SITE        126    128       CELL ATTACHMENT SITE (POTENTIAL).
FT   DISULFID    168    216       PROBABLE.
FT   DISULFID    261    309       PROBABLE.
FT   DISULFID    353    393       PROBABLE.
FT   CARBOHYD     61     61       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    103    103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    110    110       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    198    198       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    267    267       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    302    302       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    386    386       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    419    419       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   424 AA;  47676 MW;  0B7726090B31B92E CRC64;

  Query Match             7.2%;  Score 82.5;  DB 1;  Length 424;
  Best Local Similarity   22.3%;  Pred. No. 8.5;
  Matches   40;  Conservative   24;  Mismatches   60;  Indels   55;  Gaps    7;

Qy         56 ENEKALLNCYPNAKNIAGF----------FHLETDYVGLGIDRQM---------------- 90
              |:  ||  :   :|: |:           :|  | ||   :| |:
Db         47 EGKDVLLLVHNLPQNLTGYIWYKGQMTDLYHYITSYV---VDGQIIYGPAYSGRETVYSN 103

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
              | | : |    ||||  |: :|| |   ||          :  |        | | ::: |
Db        104 ASLLIQNVTQEDAGS-YTLHIIKRGDGTGG---------VTGYFTVTLYSETPKRSISSS 153

Qy        151 EVLPKSTRDAVNY----------------GMVLSVIACIQHLAKNQKIYLCGGDAKYLS 193
               : |:   :||                 |  | :   :|    |: :||  |  ||::
Db        154 NLNPREVMEAVRLICDPETPDASYLWLLNGQNLPMTHRLQLSKTNRTLYLF-GVTKYIA 211


  ESULT 13
  PL_HAEIN
ID   NPL_HAEIN      STANDARD;      PRT;   293 AA.
AC   P44539;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable N-acetylneuraminate lyase subunit (EC 4.1.3.3) (N-
DE   acetylneuraminic acid aldolase) (N-acetylneuraminate pyruvate lyase)
DE   (Nalase).
GN   NANA OR HI0142.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: N-acetylneuraminate = N-acetyl-D-mannosamine +
CC       pyruvate.
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE DHDPS FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32700; AAC21814.1; -.
DR   HSSP; P06995; 1NAL.
DR   TIGR; HI0142; -.
DR   InterPro; IPR002220; DHDPS.
DR   InterPro; IPR005264; NanA.
DR   Pfam; PF00701; DHDPS; 1.
DR   PRINTS; PR00146; DHPICSNTHASE.
DR   ProDom; PD001859; DHDPS; 1.
DR   TIGRFAMs; TIGR00683; nanA; 1.
DR   PROSITE; PS00665; DHDPS_1; 1.
DR   PROSITE; PS00666; DHDPS_2; 1.
KW   Lyase; Complete proteome.
FT   ACT_SITE    164    164       BY SIMILARITY.
SQ   SEQUENCE   293 AA;  32564 MW;  ACA755D0A5D93D33 CRC64;

  Query Match             7.1%;  Score 81.5;  DB 1;  Length 293;
  Best Local Similarity   20.8%;  Pred. No. 6.8;
  Matches   48;  Conservative   37;  Mismatches   79;  Indels   67;  Gaps   10;

Qy          6 SFTDLK---NLVLCDIGNTRIHFA--------QNYQLFSSAKEDLKRLGIQKEI--FYIS 52
              || ::|   : :: : ||  | ::           : |    :: | ||::     ||:
Db        114 SFPEIKHYYDTIIAETGNNMIVYSIPFLTGVNMGIEQFGELYKNPKVLGVKFTAGDFYLL 173

Qy         53 VNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLI 112
                 : |   |||    ||| :      ||:|          | :   ||   :: :
Db        174 ------ERLKKAYPNHLIWAGFDEMMLPAASLGVD----------GAI---GSTFNVNGV 214

Qy        113 KEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIAC 172
              :              |: |    |: |: |       :||:        |   ::  ::|
Db        215 R------------ARQIFELTKAGKLAE-------ALEI------QHVTNDLIEGILAN 248

Qy        173 IQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              :|   : : | | || |         |:| :     |  | ||  |: |
Db        249 GLYLTIKELLKLEGVDAGY---------CREPMTSKATEEQLAKAKDLKAK 290


RESULT 14
ARCC_CLOPE
ID   ARCC_CLOPE     STANDARD;      PRT;   314 AA.
AC   Q46171;
DT   01-NOV-1997 (Rel. 35, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Carbamate kinase (EC 2.7.2.2).
GN   ARCC OR CPE0171.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   MEDLINE=97199138; PubMed=9053381;
RA   Ohtani K., Bando M., Swe T., Banu S., Oe M., Hayashi H., Shimizu T.;
RT   "Collagenase gene (colA) is located in the 3'-flanking region of the
```

```
RT   perfringolysin O (pfoA) locus in Clostridium perfringens.";
RL   FEMS Microbiol. Lett. 146:155-159(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + NH(3) + CO(2) = ADP + carbamoyl
CC       phosphate.
CC   -!- PATHWAY: Arginine degradation via arginine deiminase; third step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE CARBAMATE KINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X97768; CAA66367.1; -.
DR   EMBL; AP003185; BAB79877.1; -.
DR   HSSP; P95474; 1E19.
DR   InterPro; IPR001048; Aa_kinase.
DR   InterPro; IPR003964; Bac_carb_kinase.
DR   Pfam; PF00696; aakinase; 1.
DR   PRINTS; PR01469; CARBMTKINASE.
DR   TIGRFAMs; TIGR00746; arcC; 1.
KW   Transferase; Kinase; Arginine metabolism; Complete proteome.
FT   CONFLICT      9      9       G -> E (IN REF. 1).
FT   CONFLICT    126    126       K -> T (IN REF. 1).
SQ   SEQUENCE   314 AA;  33864 MW;  170C59630EBD1A37 CRC64;

  Query Match             7.1%;  Score 81.5;  DB 1;  Length 314;
  Best Local Similarity   23.2%;  Pred. No. 7.4;
  Matches   38;  Conservative   32;  Mismatches   55;  Indels   39;  Gaps    7;

Qy         29 YQLFSSAKEDLKRLGIQKEIFYIS---VNEENEKALLNCYPNAKNIAGFF------HLET 79
              | | :: :|:| : ||:| :  |:   : ::|:    |     | |  |:       ||
Db         90 YHLQNTIREELLKRGIEKSVDTITTQVIVDKNDPGFKN---PTKPIGSSFYTKEEAEKLEK 146

Qy         80 DY-------VGLG-------------IDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLG 119
              |         | |              ::::     |::| :| |     | ::::|  |
           147 DKGYTMKEDAGRGYRRVVASPKPVDIVEKEAIKTMVDSGFIVIACGGGGIPVVEDGDRLE 206

Qy        120 GCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNY 163
              |  :| :     | :| |:||:       |:| :|    |  ||:
Db        207 G--VPAVIDKDFAAEKLAEILD-----ADALLILTAVDRVCVNF 243


RESULT 15
ACC8_HUMAN
ID   ACC8_HUMAN     STANDARD;      PRT;  1580 AA.
AC   Q09428; O75948; Q16583;
DT   01-NOV-1995 (Rel. 32, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Sulfonylurea receptor 1.
GN   ABCC8 OR SUR1 OR SUR.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Pancreatic islets;
RA   Gonzalez G., Aguilar-Bryan L., Bryan J.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Pancreas;
RA   Nishimura M., Miki T., Aizawa T., Seino S.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Foreskin, and Brain;
RA   Thomas P.T., Wohllk N., Huang E., Gagel R.F., Cote G.J.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1186-1580 FROM N.A.
RC   TISSUE=Pancreatic islets;
RX   MEDLINE=95232533; PubMed=7716548;
RA   Thomas P.M., Cote G.J., Wohllk N., Haddad B., Mathew P.M., Rabl W.,
RA   Aguilar-Bryan L., Gagel R.F., Bryan J.;
RT   "Mutations in the sulfonylurea receptor gene in familial persistent
RT   hyperinsulinemic hypoglycemia of infancy.";
RL   Science 268:426-429(1995).
RN   [5]
RP   TOPOLOGY.
RX   MEDLINE=99436117; PubMed=10506167;
RA   Raab-Graham K.F., Cirilo L.J., Boettcher A.A., Radeke C.M.,
RA   Vandenberg C.A.;
RT   "Membrane topology of the amino-terminal region of the sulfonylurea
RT   receptor.";
RL   J. Biol. Chem. 274:29122-29129(1999).
RN   [6]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=99268411; PubMed=10338089;
RA   Meissner T., Beinbrech B., Mayatepek E.;
RT   "Congenital hyperinsulinism: molecular basis of a heterogeneous
RT   disease.";
RL   Hum. Mutat. 13:351-361(1999).
RN   [7]
RP   VARIANT PHHI VAL-715.
RX   MEDLINE=96354544; PubMed=8751851;
RA   Thomas P.M., Wohllk N., Huang E., Kuhnle U., Rabl W., Gagel R.F.,
RA   Cote G.J.;
RT   "Inactivation of the first nucleotide-binding fold of the
RT   sulfonylurea receptor, and familial persistent hyperinsulinemic
RT   hypoglycemia of infancy.";
RL   Am. J. Hum. Genet. 59:510-518(1996).
RN   [8]
RP   VARIANT ALA-1368.
RX   MEDLINE=96220088; PubMed=8635661;
RA   Inoue H., Ferrer J., Welling C.M., Elbein S.C., Hoffman M.,
RA   Mayorga R., Warren-Perry M., Zhang Y., Millns H., Turner R.,
RA   Province M., Bryan J., Permutt M.A., Aguilar-Bryan L.;
RT   "Sequence variants in the sulfonylurea receptor (SUR) gene are
RT   associated with NIDDM in Caucasians.";
RL   Diabetes 45:825-831(1996).
RN   [9]
RP   VARIANT PHHI PHE-1386 DEL, AND VARIANTS GLY-1359; ALA-1368 AND
RP   ILE-1571.
RX   MEDLINE=97081768; PubMed=8923011;
RA   Nestorowicz A., Wilson B.A., Schoor K.P., Inoue H., Glaser B.,
RA   Landau H., Stanley C.A., Thornton P.S., Clement J.P. IV, Bryan J.,
RA   Aguilar-Bryan L., Permutt M.A.;
RT   "Mutations in the sulfonylurea receptor gene are associated with
RT   familial hyperinsulinism in Ashkenazi Jews.";
RL   Hum. Mol. Genet. 5:1813-1822(1996).
RN   [10]
RP   VARIANT PHHI ARG-1477.
RX   MEDLINE=96256476; PubMed=8650576;
RA   Nichols C.G., Shyng S.-L., Nestorowicz A., Glaser B., Clement J.P. IV,
RA   Gonzalez G., Aguilar-Bryan L., Permutt M.A., Bryan J.;
RT   "Adenosine diphosphate as an intracellular regulator of insulin
RT   secretion.";
RL   Science 272:1785-1787(1996).
RN   [11]
RP   VARIANTS GLN-274; MET-559; ASN-809; CYS-833 AND ALA-1368.
```

```
RX   MEDLINE=98178572; PubMed=9519757;
RA   Ohta Y., Tanizawa Y., Inoue H., Hosaka T., Ueda K., Matsutani A.,
RA   Repunte V.P., Yamada M., Kurachi Y., Bryan J., Aguilar-Bryan L.,
RA   Permutt M.A., Oka Y.;
RT   "Identification and functional analysis of sulfonylurea receptor 1
RT   variants in Japanese patients with NIDDM.";
RL   Diabetes 47:476-481(1998).
RN   [12]
RP   VARIANTS PHHI.
RX   MEDLINE=98282238; PubMed=9618169;
RA   Nestorowicz A., Glaser B., Wilson B.A., Shyng S.-L., Nichols C.G.,
RA   Stanley C.A., Thornton P.S., Permutt M.A.;
RT   "Genetic heterogeneity in familial hyperinsulinism.";
RL   Hum. Mol. Genet. 7:1119-1128(1998).
RN   [13]
RP   VARIANTS PHHI PRO-1351; CYS-1419 AND TRP-1492.
RX   MEDLINE=98443218; PubMed=9769320;
RA   Verkarre V., Fournet J.-C., de Lonlay P., Gross-Morand M.-S.,
RA   Devillers M., Rahier J., Brunelle F., Robert J.-J., Nihoul-Fekete C.,
RA   Saudubray J.-M., Junien C.;
RT   "Paternal mutation of the sulfonylurea receptor (SUR1) gene and
RT   maternal loss of 11p15 imprinted genes lead to persistent
RT   hyperinsulinism in focal adenomatous hyperplasia.";
RL   J. Clin. Invest. 102:1286-1291(1998).
RN   [14]
RP   VARIANT PHHI ASP-186.
RX   MEDLINE=99265491; PubMed=10334322;
RA   Otonkoski T., Aemmaelae C., Huopio H., Cote G.J., Chapman J.,
RA   Cosgrove K., Ashfield R., Huang E., Komulainen J., Ashcroft F.M.,
RA   Dunne M.J., Kere J., Thomas P.M.;
RT   "A point mutation inactivating the sulfonylurea receptor causes the
RT   severe form of persistent hyperinsulinemic hypoglycemia of infancy in
RT   Finland.";
RL   Diabetes 48:408-415(1999).
CC   -!- FUNCTION: PUTATIVE SUBUNIT OF THE BETA-CELL ATP-SENSITIVE
CC       POTASSIUM CHANNEL (KATP). REGULATOR OF ATP-SENSITIVE K+ CHANNELS
CC       AND INSULIN RELEASE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC       produced by alternative splicing.
CC   -!- DISEASE: DEFECTS IN ABCC8 ARE A CAUSE OF FAMILIAL PERSISTENT
CC       HYPERINSULINEMIC HYPOGLYCEMIA OF INFANCY (PHHI), AN AUTOSOMAL
CC       RECESSIVE DISORDER CHARACTERIZED BY UNREGULATED AND HIGH INSULIN
CC       SECRETION. IT CAUSES NESIDIOBLASTOSIS, A DIFFUSE ABNORMALITY OF
CC       THE PANCREAS IN WHICH THERE IS EXTENSIVE, OFTEN DISORGANIZED
CC       FORMATION OF NEW ISLETS. PHHI HAS A VARIABLE CLINICAL PHENOTYPE
CC       BUT USUALLY PRESENTS DURING THE NEONATAL PERIOD OR INFANCY WITH
CC       SEIZURES, COMA, AND OFTEN LARGE BIRTH WEIGHT FOR GESTATIONAL AGE.
CC   -!- DISEASE: DEFECTS IN ABCC8 MAY CONTRIBUTE TO AUTOSOMAL DOMINANT
CC       NON-INSULIN-DEPENDENT DIABETES MELLITUS TYPE II (MODY2 OR MODY-II)
CC       (ALSO KNOWN AS NIDDM), WHICH IS CHARACTERIZED BY AN AUTOSOMAL
CC       DOMINANT MODE OF INHERITANCE, ONSET DURING CHILDHOOD (USUALLY
CC       BEFORE 25 YEARS OF AGE) AND A PRIMARY DEFECT IN INSULIN SECRETION.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L78243; AAB02417.1; -.
DR   EMBL; L78208; AAB02417.1; JOINED.
DR   EMBL; L78209; AAB02417.1; JOINED.
DR   EMBL; L78210; AAB02417.1; JOINED.
DR   EMBL; L78211; AAB02417.1; JOINED.
DR   EMBL; L78212; AAB02417.1; JOINED.
DR   EMBL; L78255; AAB02417.1; JOINED.
DR   EMBL; L78213; AAB02417.1; JOINED.
DR   EMBL; L78214; AAB02417.1; JOINED.
DR   EMBL; L78215; AAB02417.1; JOINED.
DR   EMBL; L78216; AAB02417.1; JOINED.
DR   EMBL; L78217; AAB02417.1; JOINED.
DR   EMBL; L78218; AAB02417.1; JOINED.
DR   EMBL; L78219; AAB02417.1; JOINED.
DR   EMBL; L78220; AAB02417.1; JOINED.
DR   EMBL; L78221; AAB02417.1; JOINED.
DR   EMBL; L78222; AAB02417.1; JOINED.
DR   EMBL; L78223; AAB02417.1; JOINED.
DR   EMBL; L78225; AAB02417.1; JOINED.
DR   EMBL; L78254; AAB02417.1; JOINED.
DR   EMBL; L78226; AAB02417.1; JOINED.
DR   EMBL; L78227; AAB02417.1; JOINED.
DR   EMBL; L78228; AAB02417.1; JOINED.
DR   EMBL; L78229; AAB02417.1; JOINED.
DR   EMBL; L78230; AAB02417.1; JOINED.
DR   EMBL; L78231; AAB02417.1; JOINED.
DR   EMBL; L78232; AAB02417.1; JOINED.
DR   EMBL; L78233; AAB02417.1; JOINED.
DR   EMBL; L78234; AAB02417.1; JOINED.
DR   EMBL; L78235; AAB02417.1; JOINED.
DR   EMBL; L78236; AAB02417.1; JOINED.
DR   EMBL; L78237; AAB02417.1; JOINED.
DR   EMBL; L78238; AAB02417.1; JOINED.
DR   EMBL; L78239; AAB02417.1; JOINED.
DR   EMBL; L78240; AAB02417.1; JOINED.
DR   EMBL; L78241; AAB02417.1; JOINED.
DR   EMBL; L78242; AAB02417.1; JOINED.
DR   EMBL; L78243; AAB02418.1; -.
DR   EMBL; L78208; AAB02418.1; JOINED.
DR   EMBL; L78209; AAB02418.1; JOINED.
DR   EMBL; L78210; AAB02418.1; JOINED.
DR   EMBL; L78211; AAB02418.1; JOINED.
DR   EMBL; L78212; AAB02418.1; JOINED.
DR   EMBL; L78255; AAB02418.1; JOINED.
DR   EMBL; L78213; AAB02418.1; JOINED.
DR   EMBL; L78214; AAB02418.1; JOINED.
DR   EMBL; L78215; AAB02418.1; JOINED.
DR   EMBL; L78216; AAB02418.1; JOINED.
DR   EMBL; L78217; AAB02418.1; JOINED.

  Query Match             7.0%;  Score 81;  DB 1;  Length 1580;
  Best Local Similarity   21.1%;  Pred. No. 51;
  Matches   52;  Conservative   43;  Mismatches   85;  Indels   66;  Gaps   13;

Qy         24 HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNC---------------YPNA 68
              :| | |   |  |  ||::|    ::  :|   |  : |                | ::
Db       1174 YFIQKY--FRVASRDLQQLDDTTQLPLLSHFAETVEGLTTIRAFRYEARFQQKLLEYTDS 1231

Qy         69 KNIAGFF--------HLETDYVGLGIDRQMACLAVNN--------GVVVDAGSAITIDLI 112
               |||  |          :  :|:|  :    |  :::|         |:|  |  :|  |:
Db       1232 NNIASLFLTAANRWLEVRMEYIGACVVLIAAVTSISNSLHRELSAGLV---GLGLTYALM 1288

Qy        113 KEG------KHLGGCILP-GLAQYIHAYKKSAKILEQPPFKALDSLEVLPKSTRD------ 159
                  ::|    |  |  : ||   |:     : :: | :  ::||:  |
Db       1289 VSNYLNWMVRNLADMELQLGAVKRIHGLLKTE---AESYEGLLAPSLIPKNWPDQGKIQI 1345

Qy        160 ---AVNYGMVLSVIACIQH----LAKNQKIYLCG--GDAK---YLSAFLPHSVCKERLVF 207
                :| |   |  : ::|    ::  ||| :||  |  |    |: |      :  ::
Db       1346 QNLSVRYDSSLKPV--LKHVNALISPGQKIGICGRTGSGKSSFSLAFFRMVDTFEGHIII 1403

Qy        208 DGMEIA 213
               ||::||
Db       1404 DGIDIA 1409


Search completed: June 24, 2003, 22:12:00
Job time : 5.48585 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 9.1998 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-67
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	2	F64627	hypothetical prote
2	1115	96.7	223	2	G71887	hypothetical prote
3	391.5	34.0	209	2	H81382	hypothetical prote
4	165	14.3	246	2	D72320	conserved hypothet
5	159.5	13.8	248	2	H83111	hypothetical prote
6	158	13.7	276	2	AI2292	hypothetical prote
7	141.5	12.3	257	2	S75559	hypothetical prote
8	138.5	12.0	229	2	E70465	hypothetical prote
9	117.5	10.2	273	2	E97293	probable transcrip
10	116.5	10.1	592	2	B81009	BirA protein/Bvg a
11	116.5	10.1	592	2	H82031	probable biotin-[a
12	116	10.1	224	2	A99571	conserved hypothet
13	115.5	10.0	242	2	A82637	conserved hypothet
14	109.5	9.5	233	2	S66100	conserved hypothet
15	94	8.2	407	2	F69805	hypothetical prote
16	93.5	8.1	1164	2	C86238	protein T10O24.5 [
17	91	7.9	254	2	F83660	hypothetical prote
18	90.5	7.8	274	2	H86937	conserved hypothet
19	90.5	7.8	482	2	F72278	conserved hypothet
20	89.5	7.8	232	2	H69315	cytochrome-c3 hydr
21	89.5	7.8	272	2	A70955	hypothetical prote
22	89.5	7.8	1582	2	A56248	sulfonylurea recep
23	89	7.7	243	2	T30671	hypothetical prote
24	88.5	7.7	262	2	F70165	conserved hypothet
25	88.5	7.7	1182	2	D71638	DNA polymerase III
26	87.5	7.6	344	2	A64103	DNA-directed DNA p
27	87.5	7.6	483	2	T03204	hypothetical prote
28	86.5	7.5	393	2	AD2219	hypothetical prote
29	86	7.5	474	2	B82923	pyruvate kinase UU
30	86	7.5	627	2	AH2851	hypothetical prote
31	86	7.5	702	2	F97628	ATP-binding protei
32	85.5	7.4	240	2	AI0941	probable GntR-fami
33	85.5	7.4	401	2	E83720	methyl-accepting c
34	85.5	7.4	426	2	C64086	glycerol-3-phospha
35	84	7.3	764	2	F71863	phenylalanine-tRNA
36	83.5	7.2	450	2	AC1079	PTS system, cellob
37	83.5	7.2	452	2	AB1437	PTS system, cellob
38	83.5	7.2	839	2	B96998	phosphoenolpyruvat
39	83	7.2	303	1	D64070	ATP phosphoribosyl
40	83	7.2	553	2	T02475	hypothetical prote
41	82.5	7.2	259	2	AF1102	conserved hypothet
42	82.5	7.2	265	2	T36391	hypothetical prote
43	82.5	7.2	352	2	JS0023	glycerol-3-phospha
44	82.5	7.2	353	2	S31790	glycerol-3-phospha
45	82.5	7.2	363	2	S23137	glycerol-3-phospha

ALIGNMENTS

RESULT 1
F64627
hypothetical protein HP0862 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
C;Accession: F64627
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McK son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64627
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <TOM>
A;Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07916.1; PID:g23
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 100.0%; Score 1153; DB 2; Length 223;
Best Local Similarity 100.0%; Pred. No. 7.5e-93;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 2
G71887
hypothetical protein jhp0796 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Aug-2000
C;Accession: G71887
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric

A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: G71887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <ARN>
A;Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g415536
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0796
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 96.7%; Score 1115; DB 2; Length 223;
Best Local Similarity 94.2%; Pred. No. 1.5e-89;
Matches 210; Conservative 11; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||| |||:|:||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFKDLKDLILCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||:|||||||||||| |||:||||||||||:||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYIGLGIDRQMACLAVVNGVIVDAGSAITIDLVKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              |||||||||:||||||||||||||||||||||||:|||||||||:||:|:|||||||:|
Db        121 CILPGLAQYVHAYKKSAKILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCIQHLAKDQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
```

RESULT 3
H81382
hypothetical protein Cj0394c [imported] - Campylobacter jejuni (strain NCTC 11168)
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C;Accession: H81382
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A;Reference number: A81250; MUID:20150912; PMID:10688204
A;Accession: H81382
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-209 <PAR>
A;Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB74230.1; PID:g696786
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
C;Gene: Cj0394c
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 34.0%; Score 391.5; DB 2; Length 209;
Best Local Similarity 39.2%; Pred. No. 1.1e-26;
Matches 80; Conservative 44; Mismatches 79; Indels 1; Gaps 1;

```
Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::||||||:  :|  : : |:   :        :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
              :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203
```

RESULT 4
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 14.3%; Score 165; DB 2; Length 246;
Best Local Similarity 23.2%; Pred. No. 6.6e-07;
Matches 59; Conservative 43; Mismatches 78; Indels 74; Gaps 11;

```
Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:          |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 ----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                       :: |:||   |  :|      |   | || :          :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|       ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

Qy        144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
               |  |   |: | | : :  |:|        :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
               | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239
```

RESULT 5
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; I
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

Query Match 13.8%; Score 159.5; DB 2; Length 248;
Best Local Similarity 28.6%; Pred. No. 2e-06;
Matches 58; Conservative 34; Mismatches 60; Indels 51; Gaps 8;

```
Qy          51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
               :||  | | : |      :|       : | :||  :   ||  ||:|| :| :| ::
Db          53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy          98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139
                  :|:| |:|:| ||:  :| |||| | ||:              :|  |  : |
Db         113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy         140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
                 || :|          | :||  | :| :         :|| : |  : :|:| ||||:
Db         173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy         191 YLSAFLPHSVCKERLVFDGMEIA 213
               :   |  :     ||| |: :|
Db         222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 6
I2292
ypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

  Query Match             13.7%;  Score 158;  DB 2;  Length 276;
  Best Local Similarity   25.1%;  Pred. No. 3.1e-06;
  Matches   70;  Conservative   51;  Mismatches   94;  Indels   64;  Gaps   13;

Qy           1 MPARQSFTDLKNLVLC-DIGNTRIHFA-----------------QNYQLFSSAKEDLKR 41
               | ::||   ::|  |  :|||:|:|:|                 |   : :  |:
Db           1 MISQQSAKHIENQWLALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEV 60

Qy          42 LGIQKEIF--------------YISVNEENEKALLNCYPNAKNIA-GFFHLETDYVGLG 85
                  ||||                :|:     :  |   | | : |      |   |  ||
Db          61 GSEEKEIFFTFFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPTLG 120

Qy          86 IDRQMACLAVNNG-----VVVDAGSAITIDLIKEGKHL-GGCILPGLA-QYIHAYKKSAK 138
               |||  :|            :|:|||:|:|      ||:| || ||||:  |:    ::: :
Db         121 IDRALALWGAGMSWGFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQPASLGQQTGQ 180

Qy         139 ILEQPFKALDSLEVLPK---STRDAVNYGMVLSVIACIQ-------HLAKNQKIYLCGGD 188
               : :   :|: ||   |:   :| :|:  |:: ::|| ::        |  : |: : |||
Db         181 LPQVEMEAIKSLP--PRFALNTTEAIQSGVIYTLIAGMRDFTEEWLSLFPDGKVAIKGGD 238

Qy         189 ----AKYLSAFLP----HSVCKERLVFDGMEIALKKAGI 219
                   || |  |      : :  |:| ||:  :  ||:
Db         239 RILLLNYLQALYPDLAARLIVEPNLIFWGMQTIV--AGV 275


RESULT 7
S75559
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75559
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocy
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75559
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-257 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18120.1; PID:d10
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

  Query Match             12.3%;  Score 141.5;  DB 2;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 7.6e-05;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy          18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
               |||:|:|:|     : |    |:                ||    :    | |: |
Db          25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy          59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                         |:| ||:                |||| :| |          :|||
Db          80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy         104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
               |:|:||    : | | || |||||   :       ::   |   :| |  ||      |
Db         124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy         159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
                |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||        : |
Db         182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

Qy         206 VFDGM 210
               :| ||
Db         242 IFLGM 246


RESULT 8
E70465
hypothetical protein aq_1924 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
C;Accession: E70465
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: E70465
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-229 <AQF>
A;Cross-references: GB:AE000763; NID:g2984178; PIDN:AAC07720.1; PID:g2984188; GB:AE0
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_1924

  Query Match             12.0%;  Score 138.5;  DB 2;  Length 229;
  Best Local Similarity   26.5%;  Pred. No. 0.00012;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy          17 DIGNTRIHFA-------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
               |:||: :  |              : :        || ||       ||| :   : :
Db           7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy          64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                 |  | :    | :: ||     || ||      :       | ||: ||:|: |||: |
Db          60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy         115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
               ||  || |  ||         |  |||    : :    : :|: | :|||: |  |
Db         120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172
```

```
Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219
```

RESULT 9
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
C;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

```
  Query Match             10.2%;  Score 117.5;  DB 2;  Length 273;
  Best Local Similarity   21.4%;  Pred. No. 0.01;
  Matches   57;  Conservative   48;  Mismatches   94;  Indels   67;  Gaps   11;

Qy         10 LKNLVLCDIGNTRI--------HFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKAL 61
Db         10 LKVILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQ--VMNLFQQDKLDPTL 67

Qy         62 LN-----------CYPNAKNIAGFFHLETDYVGLGI----------------DRQMACLA 94
Db         68 VEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVA 127

Qy         95 VN-----NGVVVDAGSAITIDLIKE-GKHLGGCILPGLAQYIHA-YKKSAKI----LEQP 143
Db        128 AHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKP 187

Qy        144 FKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQK------------IYLCGGDAKY 191
Db        188 AYA------ICKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKL 241

Qy        192 LSAFLPH-SVCKERLVFDGMEIALKK 216
Db        242 ISEEAKNVDVINPFLTLEGLRIIYEK 267
```

RESULT 10
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain MC
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

```
  Query Match             10.1%;  Score 116.5;  DB 2;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.032;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
Db        551 AKVAEALPPAFLAENTV 567
```

RESULT 11
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidi
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C;Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

```
  Query Match             10.1%;  Score 116.5;  DB 2;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.032;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
Db        551 AKVAEALPPAFLAENTV 567
```

RESULT 12

```
A99571
conserved hypothetical protein. MYPU_4730 [imported] - Mycoplasma pulmonis (strain UAB C
C;Species: Mycoplasma pulmonis
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C;Accession: A99571
R;Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A;Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A;Reference number: A99512; MUID:21267165; PMID:11353084
A;Accession: A99571
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-224 <KUR>
A;Cross-references: GB:AL445566; PID:g14089887; PIDN:CAC13646.1; GSPDB:GN00153
A;Experimental source: strain UAB CTIP
C;Genetics:
A;Gene: MYPU_4730
A;Genetic code: SGC3

  Query Match            10.1%;  Score 116;  DB 2;  Length 224;
  Best Local Similarity  22.3%;  Pred. No. 0.011;
  Matches   51;  Conservative   30;  Mismatches   94;  Indels   54;  Gaps    7;

Qy         14 VLCDIGNTRIHFA-----------------------QNYQLFSSAKEDLKRLGIQKEI 48
                :  |:||:   ||                       ::: ||   | |:     |||
Db          3 IFIDLGNSFAKFALKNENKAHFLFRLKTQSVVDPSFEIKSFNLFEFNKLDV------KEI 56

Qy         49 FYISVNEENEKALL-----NCYPNAKNIAGFFHLE-----------TDYVGLGI--DRQM 90
               ||    |  :|     : : ||| |  | |              |  :|| |  :
Db         57 LICSVRNAKENQILEFKLKSIFKNAK-IDFFIHKSQSLVKFCQKELTSEIGLDIVANAYY 115

Qy         91 ACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSL 150
                | : :  |:|| |  ||     |  | ||: |    :      :|  |  :
Db        116 VLHKSKNAIFISLGTAIVITQIKNSSIEGISIYPGIYQSFKNFFNVVAKIESNFNIIKIP 175

Qy        151 EVLPKSTRDAVN----YGMVLSVIACIQHLAKNQKIYLCGGDAKYLSAF 195
               :| |:| ::::     | |  :   |  : |   |:: ||   :   |
Db        176 PILGKTTLESISSSLVRGSVFLLKGYIDEIDKTSDIFITGGTITFSKIF 224


RESULT 13
A82637
conserved hypothetical protein XF1795 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 ature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84603.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1795

  Query Match            10.0%;  Score 115.5;  DB 2;  Length 242;
  Best Local Similarity  31.5%;  Pred. No. 0.013;
  Matches   45;  Conservative   22;  Mismatches   55;  Indels   21;  Gaps    6;

Qy         85 GIDRQMACL---AVNNGVVVDAGSAITIDLI-KEGKHLGGCILPG---LAQYIHAYKKSA 137
               |:|| :| |       | :||  |:|:||||:   | |||| |      : | :|     |
Db         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152

Qy        138 KILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACI--------QHLAKNQKIYLCGGDA 189
               :  : |    : ||   : | ||:  |   : :| |        | | :: :: | ||
Db        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211

Qy        190 KYLSAFLPHSVCKERLVFDGMEI 212
                 |  :|   | :  || ||: |
Db        212 ASLLPWLGDVVHRPTLVLDGLAI 234


RESULT 14
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S66100; E69740
R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis c
A;Reference number: S65967; MUID:96051385; PMID:7584024
A;Accession: S66100
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <OGA>
A;Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g46
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 199
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e1182
A;Experimental source: strain 168
C;Genetics:
A;Gene: yacB
A;Start codon: TTG
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match            9.5%;  Score 109.5;  DB 2;  Length 233;
  Best Local Similarity  26.7%;  Pred. No. 0.041;
  Matches   35;  Conservative   24;  Mismatches   47;  Indels   25;  Gaps    5;

Qy         74 FFHLETDYVG----------------LGIDRQMACLAV-----NNGVVVDAGSAITIDLI 112
               :||:|   ||                 :| || :  :|      |  :||| |:| |   |
Db         80 YFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYI 139

Qy        113 KEGK-HLGGCILPGLAQYIHA-YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVI 170
               | | .::|| | ||:    | | ::||:        |:  :: |:|  |:  |::   :
Db        140 DENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDN--IIGKNTVSAMQSGILFGYV 197
```

```
Qy        171 ACIQHLAKNQK 181
                :: : |  |
Db        198 GQVEGIVKRMK 208


RESULT 15
F69805
hypothetical protein yfjB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: F69805
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
euchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: F69805
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-407 <KUN>
A;Cross-references: GB:Z99108; GB:AL009126; NID:g2633055; PIDN:CAB12645.1; PID:g2633140
A;Experimental source: strain 168
C;Genetics:
A;Gene: yfjB
C;Superfamily: Bacillus subtilis hypothetical protein yfjB

  Query Match             8.2%;  Score 94;  DB 2;  Length 407;
  Best Local Similarity   21.8%;  Pred. No. 1.8;
  Matches   52;  Conservative   25;  Mismatches   70;  Indels   92;  Gaps    7;

Qy         15 LCDIGNTRIHF-----------------------AQNYQLFSSAKEDLKRLGIQKEIFY 50
              | |:| | : |                      |:|| :    | |||
Db        213 LKDVGKTGLRFAKGNAITATLITGATETFGAGVKIAENYAKYQDKPEVLKR--------- 263

Qy         51 ISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITID 110
                 || ||: |   |   :||              |   :   :   ||   |:
Db        264 -----ENAKAVGNAVNNTIFVAG-------------GATAGAVIGGAIGSFAGPVGTVL 304

          111 LIKEGKHLGGCILPGLAQYIHAY-KKSAKILEQPF--------KALDSLEVLPKSTRDAV 161
              |   | ::|| :   :|:|   : :|:|  |::|         | |:|   : ||  | :
Db        305 LGAAGSYIGGVVGEQIAKYTAGFAEKAALKLKEPIHAVVDTAKKGLESAGKVVKSVNDGI 364

Qy        162 NYGMVLSVIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGIL 220
              :              | |: |:     ||                  ||| | | |  |
Db        365 D-------------AANKSIHKGIDSAK----------------KGMEKAKKTADSL 392


Search completed: June 24, 2003, 22:19:26
Job time : 9.1998 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 25.2424 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-67
Perfect score: 1153
Sequence: 1 MPARQSFTDLKNLVLCDIGN.....RLVFDGMEIALKKAGILECK 223

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1153	100.0	223	19	AAW98422	H. pylori GHPO 344
2	1153	100.0	223	23	AAU91161	Helicobacter pylor
3	1153	100.0	223	23	AAU91181	Pantothenate kinas
4	391.5	34.0	209	23	AAU91165	Pantothenate kinas
5	165	14.3	246	23	AAU91156	Thermotoga maritim
6	159.5	13.8	248	23	AAU91164	Pantothenate kinas
7	148	12.8	249	23	AAU91182	Pantothenate kinas
8	144.5	12.5	241	23	AAU91179	Pantothenate kinas
9	144.5	12.5	249	23	AAU91178	Pantothenate kinas
10	141.5	12.3	257	23	AAU91160	Synechocystis pant
11	138.5	12.0	229	23	AAU91159	Aquifex aeolicus p
12	129.5	11.2	244	23	AAU91168	Pantothenate kinas
13	116.5	10.1	455	21	AAY74910	Neisseria meningit
14	116.5	10.1	592	20	AAY38615	Neisseria meningit
15	116.5	10.1	592	20	AAY38616	Neisseria meningit
16	116.5	10.1	592	21	AAY74912	Neisseria meningit
17	116.5	10.1	592	21	AAY74913	Neisseria meningit
18	116.5	10.1	592	23	AAU91166	Pantothenate kinas
19	116.5	10.1	592	23	AAU91169	Pantothenate kinas
20	116	10.1	455	20	AAY38617	Neisseria gonorrho
21	116	10.1	455	21	AAY74908	Neisseria gonorrhe
22	116	10.1	460	23	AAU91167	Pantothenate kinas
23	116	10.1	592	20	AAY38618	Neisseria gonorrho
24	116	10.1	592	21	AAY74911	Neisseria gonorrhe
25	115.5	10.0	242	23	AAU91180	Pantothenate kinas
26	111.5	9.7	389	21	AAY74909	Neisseria meningit
27	109.5	9.5	233	23	AAU91163	Pantothenate kinas
28	109.5	9.5	258	22	AAU01243	B. subtilis novel
29	109.5	9.5	258	23	AAU91149	Bacillus subtilis
30	108	9.4	219	23	AAU91176	Pantothenate kinas
31	105.5	9.2	257	23	AAU91174	Pantothenate kinas
32	102.5	8.9	250	23	AAU91150	Clostridium acetob
33	102	8.8	212	23	AAU91177	Pantothenate kinas
34	101	8.8	256	23	AAU91175	Pantothenate kinas
35	100.5	8.7	255	23	AAU91154	Geobacter sulfurre
36	98.5	8.5	262	23	AAU91170	Pantothenate kinas
37	94.5	8.2	338	22	ABB67066	Drosophila melanog
38	91	7.9	254	23	AAU91171	Pantothenate kinas
39	89.5	7.8	272	22	AAG81225	Mycobacterium tube
40	89.5	7.8	272	23	AAU91152	Mycobacterium tube
41	89.5	7.8	1582	16	AAR77088	Hamster sulphonylu
42	89	7.7	594	22	ABB61659	Drosophila melanog
43	88.5	7.7	262	23	AAU91158	Borrelia burgdorfe
44	87.5	7.6	258	23	AAU91172	Pantothenate kinas
45	87.5	7.6	344	23	ABB84735	DNA polymerase III

ALIGNMENTS

RESULT 1
AAW98422
ID AAW98422 standard; Protein; 223 AA.
XX
AC AAW98422;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 344 protein.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
XX
OS Helicobacter pylori.
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
XX
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX
DR WPI; 1998-542293/46.
DR N-PSDB; AAX14141.
XX

PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 8; Page 709-710; 2054pp; English.
XX
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 223 AA;

Query Match 100.0%; Score 1153; DB 19; Length 223;
Best Local Similarity 100.0%; Pred. No. 4.2e-111;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
Db 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
Db 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
Db 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
Db 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223

RESULT 2
AAU91161
ID AAU91161 standard; Protein; 223 AA.
XX
AC AAU91161;
XX
DT 05-JUN-2002 (first entry)
XX
DE Helicobacter pylori pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Helicobacter pylori.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 78-79; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 223 AA;

Query Match 100.0%; Score 1153; DB 23; Length 223;
Best Local Similarity 100.0%; Pred. No. 4.2e-111;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
Db 1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
Db 61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
Db 121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
Db 181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223

RESULT 3
AAU91181
ID AAU91181 standard; Protein; 223 AA.
XX
AC AAU91181;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #19.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Helicobacter pylori.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54202.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 112-113; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate

```
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   223 AA;

  Query Match             100.0%;  Score 1153;  DB 23;  Length 223;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-111;
  Matches  223;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MPARQSFTDLKNLVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKA 60

Qy         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGG 120

Qy        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ 180

Qy        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223
              |||||||||||||||||||||||||||||||||||||||||||
Db        181 KIYLCGGDAKYLSAFLPHSVCKERLVFDGMEIALKKAGILECK 223


RESULT 4
AAU91165
ID   AAU91165 standard; Protein; 209 AA.
XX
AC   AAU91165;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #3.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Campylobacter jejuni.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 84-85; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   209 AA;

  Query Match             34.0%;  Score 391.5;  DB 23;  Length 209;
  Best Local Similarity   39.2%;  Pred. No. 2.7e-32;
  Matches   80;  Conservative   44;  Mismatches   79;  Indels    1;  Gaps    1;

Qy         13 LVLCDIGNTRIHFAQNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIA 72
              ::||||||:  :|  : : |:   :        :::||||:|||  :: | |   |  |:
Db          1 MLLCDIGNSNANFLDDNKYFTLNIDQFLEFKNEQKIFYINVNEHLKEHLKN-QKNFINLE 59

Qy         73 GFFHLETDYVGLGIDRQMACLAVNNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA 132
               :|  :| | ||||||  ||  : :|||||||||||||:|    |||| ||||:| |
Db         60 PYFLFDTIYQGLGIDRIAACYTIEDGVVVDAGSAITIDIISNSIHLGGFILPGIANYKKI 119

Qy        133 YKKSAKILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQKIYLCGGDAKYL 192
              |   :  |:  |    ||:  |: | ||::||:   :   |:  |:|:|:|  ||| ::|
Db        120 YSHISPRLKSEFNTQVSLDAFPQKTMDALSYGVFKGIYLLIKDAAQNKKLYFTGGDGQFL 179

Qy        193 SAFLPHSVCKERLVFDGMEIALKK 216
              : :  |::  : |:| ||:  :|:
Db        180 ANYFDHAIYDKLLIFRGMKKIIKE 203


RESULT 5
AAU91156
ID   AAU91156 standard; Protein; 246 AA.
XX
AC   AAU91156;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Thermotoga maritima pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Thermotoga maritima.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 74-75; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
```

```
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   246 AA;

  Query Match          14.3%;  Score 165;  DB 23;  Length 246;
  Best Local Similarity   23.2%;  Pred. No. 9.7e-09;
  Matches   59;  Conservative   43;  Mismatches   78;  Indels   74;  Gaps   11;

Qy         14 VLCDIGNT----------------RIH----------FAQNYQLFSSAKEDLKRLGI--- 44
              :| |:|||                |:           |:  : |   |  ::| :|:
Db          3 LLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHLHPLLGDAMREIKGIGVASV 62

Qy         45 ----------QKEIFYIS---VNEENEKALLNCYP-NAKNIAGFFHLETDYVGLGIDRQ 89
                      :: |:||   |  :|      |   | || :           :| ||
Db         63 VPTQNTVIERFSQKYFHISPIWVKAKN-----GCVKWNVKNPS----------EVGADRV 107

Qy         90 MACLAV-----NNGVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHA-YKKSAKILEQP 143
                 :|      ||:::| |:| |:||:  | : || ||||    :|: :: :||:
Db        108 ANVVAFVKEYGKNGIIIDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLVE 167

          144 FKALDSLEVLPKSTRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFL 196
               |  |   |: | | : :  |:|        :|  |: :  :  : | || :| :   :
Db        168 VKPADF--VVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYGDLPVVLTGGQSKIVKDMI 225

Qy        197 PHSVCKERLVFDGM 210
               | :  | |   |:
Db        226 KHEIFDEDLTIKGV 239


RESULT 6
AAU91164
ID   AAU91164 standard; Protein; 248 AA.
XX
AC   AAU91164;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #2.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
     24-AUG-2001; 2001WO-US26531.

     24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 83-84; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   248 AA;

  Query Match          13.8%;  Score 159.5;  DB 23;  Length 248;
  Best Local Similarity   28.6%;  Pred. No. 3.7e-08;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97
              :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::
Db         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112

Qy         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA-------------QYIHAYKKSAKI 139
                  :|:| |:|:| ||:  :| |||| | ||:                :|  |  : |
Db        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190
                 || :|          | :||  | :| :         :|| : |  : :|:| ||||:
Db        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221

Qy        191 YLSAFLPHSVCKERLVFDGMEIA 213
               :   |  :     ||| |:.:|
Db        222 LVRDELAGARIMPDLVFVGLALA 244


RESULT 7
AAU91182
ID   AAU91182 standard; Protein; 249 AA.
XX
AC   AAU91182;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #20.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas syringae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54203.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 113-114; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
```

```
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match            12.8%;  Score 148;  DB 23;  Length 249;
  Best Local Similarity   29.3%;  Pred. No. 5.7e-07;
  Matches   58;  Conservative   34;  Mismatches   78;  Indels   28;  Gaps    7;

Qy         34 SAKEDLKRLGIQKEIFYISVNEENEKALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACL 93
              ||:|  | :    : | ||         |   |  :||  :   |:  ||:|| :| :
Db         58 SAEETAKLVCALADTFSISP---------VCAAPAPELAGVINGYDDFARLGLDRWLAFV 108

Qy         94 AVNNGV-----VVDAGSAITIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILE----QP 143
                : |     |:| |:||| | ::  |||||| | ||:    :  :   : :     :
Db        109 GAYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEA 168

Qy        144 FKALDSLEVLPKSTRDAVNYG--MVLSVIACIQ------HLAKNQKIYLCGGDAKYLSAF 195
               |||  | |  ::| :||  |  ::|  | |       :   :  |:: |||| ::
Db        169 EKALVRL-VPGRATAEAVERGCSLMLRGFAMTQIEIAREYWGDDFAIFVTGGDAVLVADV 227

Qy        196 LPHSVCKERLVFDGMEIA 213
              || :     ||| |: :|
Db        228 LPGARIVPDLVFVGLALA 245


RESULT 8
AAU91179
ID   AAU91179 standard; Protein; 241 AA.
XX
AC   AAU91179;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #17.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Thiobacillus ferrooxidans.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54200.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 109-110; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   241 AA;

  Query Match            12.5%;  Score 144.5;  DB 23;  Length 241;
  Best Local Similarity   24.4%;  Pred. No. 1.3e-06;
  Matches   60;  Conservative   40;  Mismatches   97;  Indels   49;  Gaps   10;

Qy         13 LVLCDIGNTRIHFAQNYQ--LFSSAK--------EDLKRLGIQKEIFYISVNEENEKALL 62
              ::   :||||   |  :    | |          | |:: |:     ::|     ||
Db          1 MIFIAVGNTRTLLAHTHDGVHFDSVSVATSLPPTEILQQPGLT----WLSAPNREPVALG 56

Qy         63 NCYPNA-------------KNIAGFFH--LETDY---VGLGIDRQMACLAV------NN 97
                | |               :   |||   :  ||     || ||:   ||       :
Db         57 GVVPAALAAWREALATAEVREPDPGFFRRAVPHDYHPPESLGFDRRCCLLAAAMDYPGQD 116

Qy         98 GVVVDAGSAITIDLIKEGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDS-LEVLPKS 156
               :|:| |:|||||:  |   || ||||:|  :    :  :|  |   |:: |:|
Db        117 SIVIDMGTAITIDLLAGGHFRGGRILPGIAMSLRGLHEGTALL--PEVVLNAPAEMLGND 174

Qy        157 TRDAVNYGMV-------LSVIACIQHLAKNQKIYLCGGDAKYLSAFLPHSVCKERLVFDG 209
              | :|:  |::          |   :  :   :| : ||||:     :  |: :  |:  |
Db        175 TSNAIQAGVIHLFADALRGAITDFRQYSPQARILITGGDAERWQPGIAGSLYQPHLLLRG 234

Qy        210 MEIALK 215
                : ::
Db        235 FYLWIR 240


RESULT 9
AAU91178
ID   AAU91178 standard; Protein; 249 AA.
XX
AC   AAU91178;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #16.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas putida.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54199.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 108-109; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
```

```
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match             12.5%;  Score 144.5;  DB 23;  Length 249;
  Best Local Similarity   29.8%;  Pred. No. 1.3e-06;
  Matches   61;  Conservative   26;  Mismatches   63;  Indels   55;  Gaps   10;

Qy         51 ISVNEENEKALLNC--------------YPNAKNIAGFFHLETDYVGLGIDRQMAC---- 92
              :||  | |   | |               :|  : :||  :   ||  ||:|| :|
Db         54 VSVRSEEETDAL-CALIAQAFAVQAKVAHP-VREMAGVRNGYDDYQRLGMDRWLAALGAF 111

Qy         93 -LAVNNGVVVDAGSAITIDLIK-EGKHLGGCILPGL----------AQYIHAYKKSAKIL 140
               ||    :|:| |:|   | :  :|:|||| | ||:          : |     ||:
Db        112 HLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRIRYDDASAE-- 169

Qy        141 EQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHLAKNQ-----------KIYLCGGD 188
              :|| ||    :|| :||  | ||     :|  |  |              ::| |||
Db        170 ----RALSSLSP-GRSTVEAVERGCVL----MLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        189 AKYLSAFLPHSVCKERLVFDGMEIA 213
              |  : | || :     ||| |: :|
Db        221 APLVRAALPQARVVPDLVFVGLAMA 245


RESULT 10
AAU91160
ID   AAU91160 standard; Protein; 257 AA.
XX
AC   AAU91160;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Synechocystis pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Synechocystis sp.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
     (OMNI-) OMNIGENE BIOPRODUCTS INC.

PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 77-78; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   257 AA;

  Query Match             12.3%;  Score 141.5;  DB 23;  Length 257;
  Best Local Similarity   26.9%;  Pred. No. 2.8e-06;
  Matches   66;  Conservative   28;  Mismatches   76;  Indels   75;  Gaps   12;

Qy         18 IGNTRIHFAQNYQLFSSAKEDLKR--------------LGIQKEIFYISVNEENE---- 58
              |||:|:|:|     : |    |:                ||    :    | |: |
Db         25 IGNSRLHWA-----YCSGNAPLQTWVTDYNPKSAQLPVLLGKVPLMLASVVPEQTEVWRV 79

Qy         59 ----------KALLNCYPNAKNIAGFFHLETDYVGLGIDRQMACLAVN-----NGVVVDA 103
                        |:| ||:                |||| :| |           :|||
Db         80 YQPKILTLKNLPLVNLYPS---------------FGIDRALAGLGTGLTYGFPCLVVDG 123

Qy        104 GSAITIDLIKEGKHL-GGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK----STR 158
              |:|:||    : | | || |||||   :       ::    |   :| |  ||      |
Db        124 GTALTITGFDQDKKLVGGAILPGLG--LQLATLGDRLAALPKLEMDQLTELPDRWALDTP 181

Qy        159 DAVNYGMVLSVIACIQ-HLAKNQKIY------LCGGDAKYLSAFL-PHS-----VCKERL 205
               |:  |:|  |:  :| :|   ||::      : ||| | |  ||  ||        : |
Db        182 SAIFSGVVYGVLGALQSYLQDWQKLFPGAAMVITGGDGKILHGFLKEHSPNLSVAWDDNL 241

Qy        206 VFDGM 210
              :| ||
Db        242 IFLGM 246


RESULT 11
AAU91159
ID   AAU91159 standard; Protein; 229 AA.
XX
AC   AAU91159;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Aquifex aeolicus pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Aquifex aeolicus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 77; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
```

```
SQ   Sequence   229 AA;

  Query Match           12.0%;  Score 138.5;  DB 23;  Length 229;
  Best Local Similarity   26.5%;  Pred. No. 4.9e-06;
  Matches   61;  Conservative   32;  Mismatches   84;  Indels   53;  Gaps   11;

Qy         17 DIGNTRIHFA------------QNYQLFSSAKEDLKRLGIQKEIFYISVNEENEKALLN 63
              |:||: :  |              : :         || ||       ||| :   : :
Db          7 DVGNSSVDIALWEGKKVKDFLKLSHEEFLKEEFPKLKALG-------ISVKQSFSEKVRG 59

Qy         64 CYPNAKNI-AGFFHLETDY---VGLGIDRQMACLAV-----NNGVVVDAGSAITIDLIKE 114
                |  | :    | :: ||     || ||     :       | ||: ||:|: |||: |
Db         60 KIPKIKFLKKENFPIQVDYKTPETLGTDRVALAYSAKKFYGKNVVVISAGTALVIDLVLE 119

Qy        115 GKHLGGCILPGLAQYIHAYKKSAKILEQPFKAL-----DSLEV-LPKSTRDAVNYG---- 164
              ||  || |  ||        |  |||    : :     : :|: | :|||: |  |
Db        120 GKFKGGFITLGLG-------KKLKILSDLAEGIPEFFPEEVEIFLGRSTRECVLGGAYRE 172

Qy        165 ---MVLSVIACIQHLAKNQ-KIYLCGGDAKYLSAFLPHSVCKERLVFDGM 210
                 : | :   : : | : |: : ||: || | |    :    ||  ||
Db        173 STEFIKSTLKLWRKVFKRKFKVVITGGEGKYFSKF---GIYDPLLVHRGM 219


RESULT 12
AAU91168
ID   AAU91168 standard; Protein; 244 AA.
XX
AC   AAU91168;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #6.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54189.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 8; Page 94-95; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   244 AA;

  Query Match           11.2%;  Score 129.5;  DB 23;  Length 244;
  Best Local Similarity   26.0%;  Pred. No. 4.6e-05;
  Matches   58;  Conservative   40;  Mismatches   82;  Indels   43;  Gaps   12;

Qy         12 NLVLCDIGNT--RIHFAQNYQLFS------SAKEDLKRL----GIQKEIF-YISVNEENE 58
              ||:: | ||:  :: | :|  : |       | : |  |       | |:  : : :|
Db          4 NLIV-DQGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEA 62

Qy         59 KALLNCYPNAKNIAGF-----FHLETDYVGLGIDRQMACLAVNN------GVVVDAGSAI 107
              :|::     |  : |        |: |   || ||  | :  ::       :|:|||:||
Db         63 EAIVRSCAAASLMMGTETPVPLRLQYDRRTLGADRLAAVVGAHSLYPNTELLVIDAGTAI 122

Qy        108 TIDLIK-EGKHLGGCILPGLAQYIHAYKKSAKILEQPFKALDSLEVLPK------STRDA 160
              | : :  || :||| | |||    |   |:  :       :|   : ||      || :|
Db        123 TYERVSAEGIYLGGNISPGL----HLRFKALHLFTGRLPLIDPSGISPKIAEYGSSTEEA 178

Qy        161 VNYGMVLSVIACIQ------HLAKNQK-IYLCGGDAKYLSAFL 196
              :  |::  :   |       |  : :  : | |||| ||:  :
Db        179 ITAGVIHGLAGEIDRYIDDLHAKEGRSAVILTGGDANYLARII 221


RESULT 13
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
```

```
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             10.1%;  Score 116.5;  DB 21;  Length 455;
  Best Local Similarity   24.9%;  Pred. No. 0.0024;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        194 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 250

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |      | ||    |
Db        251 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 306

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        307 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 362

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||        | :| :||  ||: :|   :  :             |  : : || |
Db        363 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 413

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        414 AKVAEALPPAFLAENTV 430


RESULT 14
AAY38615
ID   AAY38615 standard; Protein; 592 AA.
XX
AC   AAY38615;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.

     WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             10.1%;  Score 116.5;  DB 20;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0034;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
              |   :|  | :||          :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
                ||        | :| :||  ||: :|   :  :             |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


RESULT 15
AAY38616
ID   AAY38616 standard; Protein; 592 AA.
XX
AC   AAY38616;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
```

```
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             10.1%;  Score 116.5;  DB 20;  Length 592;
  Best Local Similarity   24.9%;  Pred. No. 0.0034;
  Matches   64;  Conservative   37;  Mismatches   85;  Indels   71;  Gaps   13;

Qy          1 MPARQSFTDLKNLVLCDIGNTRIHFA--QN---YQLFSSAKEDLKRLGIQ--------KE 47
              :| |:   | :  :| | ||:|: :|  :|     : |:   ||  || :
Db        331 VPKRR---DSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVR 387

Qy         48 IFYISVNEENEKA--------LLNCYPNAKNIAG----FFHLETDYVGLGIDRQMACL-- 93
               |   :|  | :||          :   |::    |    : | |      | ||    |
Db        388 IVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH----GSDRWFNALGS 443

Qy         94 ---AVNNGVVVDAGSAITID-LIKEGKHLGGCILPGLAQYIHAYKKSAKI---------- 139
                 : |  |||  |:|:|:| |  :| :||| |:||     |  |:|  :
Db        444 RRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG----FHLMKESLAVRTANLNRHAG 499

Qy        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACIQHL----------AKNQKIYLCGGDA 189
               ||         | :| :||  ||: :|   :  :           |   : : || |
Db        500 KRYPF---------PTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGA 550

Qy        190 KYLSAFLPHSVCKERLV 206
              ::  || :   |  |
Db        551 AKVAEALPPAFLAENTV 567


Search completed: June 24, 2003, 22:08:42
Job time : 25.2424 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 8.60143 Seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453A-59
Perfect score: 1077
Sequence: 1 MTQHFLLFDIGNTNVKIGIA.....SLNHGFIFGFAAMTEGVLAA 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	84.5	7.8	486	1	US-07-872-678A-48	Sequence 48, Appl
2	79.5	7.4	531	4	US-09-360-197-14	Sequence 14, Appl
3	78.5	7.3	455	4	US-09-240-639-10	Sequence 10, Appl
4	78	7.2	1711	2	US-08-342-930-2	Sequence 2, Appli
5	75.5	7.0	462	4	US-09-129-112-15	Sequence 15, Appl
6	75	7.0	325	1	US-08-276-919-4	Sequence 4, Appli
7	75	7.0	325	1	US-08-276-919-13	Sequence 13, Appl
8	75	7.0	325	1	US-08-776-088-4	Sequence 4, Appli
9	75	7.0	325	1	US-08-776-088-18	Sequence 18, Appl
10	75	7.0	325	5	PCT-US95-09145A-4	Sequence 4, Appli
11	75	7.0	325	5	PCT-US95-09145A-18	Sequence 18, Appl
12	75	7.0	354	1	US-08-276-919-2	Sequence 2, Appli
13	75	7.0	354	1	US-08-776-088-2	Sequence 2, Appli
14	75	7.0	354	1	US-08-776-088-6	Sequence 6, Appli
15	75	7.0	354	3	US-09-325-320-2	Sequence 2, Appli
16	75	7.0	354	4	US-09-585-109-2	Sequence 2, Appli
17	75	7.0	354	5	PCT-US95-09145A-2	Sequence 2, Appli
18	75	7.0	354	5	PCT-US95-09145A-6	Sequence 6, Appli
19	74.5	6.9	485	4	US-08-818-112-70	Sequence 70, Appl
20	74.5	6.9	485	4	US-08-818-111-71	Sequence 71, Appl
21	74.5	6.9	485	4	US-09-056-556-70	Sequence 70, Appl
22	74.5	6.9	485	4	US-09-072-596-71	Sequence 71, Appl
23	73.5	6.8	933	4	US-08-764-870-14	Sequence 14, Appl
24	73.5	6.8	933	4	US-08-980-115-14	Sequence 14, Appl
25	73	6.8	463	2	US-08-679-635A-4	Sequence 4, Appli
26	73	6.8	463	4	US-09-419-163-4	Sequence 4, Appli
27	72.5	6.7	227	4	US-09-097-889-16	Sequence 16, Appl
28	72.5	6.7	490	4	US-09-215-694-14	Sequence 14, Appl
29	72.5	6.7	3010	4	US-09-014-416-3	Sequence 3, Appli
30	72	6.7	459	4	US-09-129-112-9	Sequence 9, Appli
31	71.5	6.6	274	1	US-08-776-088-15	Sequence 15, Appl
32	71.5	6.6	274	5	PCT-US95-09145A-15	Sequence 15, Appl
33	71.5	6.6	594	3	US-08-729-416C-7	Sequence 7, Appli
34	71.5	6.6	2013	1	US-08-324-977-12	Sequence 12, Appl
35	71.5	6.6	2013	2	US-08-384-616-12	Sequence 12, Appl
36	71.5	6.6	2013	2	US-08-904-686A-12	Sequence 12, Appl
37	71.5	6.6	2013	4	US-09-315-850-12	Sequence 12, Appl
38	71.5	6.6	2201	4	US-08-952-981A-2	Sequence 2, Appli
39	71.5	6.6	2620	1	US-08-324-977-32	Sequence 32, Appl
40	71.5	6.6	2620	2	US-08-384-616-32	Sequence 32, Appl
41	71.5	6.6	2620	2	US-08-904-686A-32	Sequence 32, Appl
42	71.5	6.6	2620	4	US-09-315-850-32	Sequence 32, Appl
43	71.5	6.6	2621	1	US-08-324-977-36	Sequence 36, Appl
44	71.5	6.6	2621	2	US-08-384-616-36	Sequence 36, Appl
45	71.5	6.6	2621	2	US-08-904-686A-36	Sequence 36, Appl

ALIGNMENTS

```
RESULT 1
US-07-872-678A-48
; Sequence 48, Application US/07872678A
; Patent No. 5541060
;  GENERAL INFORMATION:
;    APPLICANT:  Bell, Graeme, et al.
;    TITLE OF INVENTION:  DETECTION OF EARLY-ONSET
;    TITLE OF INVENTION:  NON-INSULIN-DEPENDENT DIABETES MELLITUS
;    NUMBER OF SEQUENCES:  48
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  Post Office Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/872,678A
;      FILING DATE:  22-APRIL-1992
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coughlin, Daniel F.
;      REGISTRATION NUMBER:  36,111
;      REFERENCE/DOCKET NUMBER:  ARCD016
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713-787-1400
;      TELEFAX:  713-789-2679
;      TELEX:  79-0924
;  INFORMATION FOR SEQ ID NO:  48:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  486 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-872-678A-48

  Query Match             7.8%;  Score 84.5;  DB 1;  Length 486;
  Best Local Similarity   20.5%;  Pred. No. 0.23;
  Matches   56;  Conservative   32;  Mismatches   68;  Indels  117;  Gaps   13;

Qy          5 FLLFDIGNTNVKI------GIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPA 58
              ||  |:| ||:::      |       : | || |  :|| :                |
Db         82 FLAIDLGGTNLRVVLVKLGGDRTFDTTQSKYRLP-DAMRTTQN--------------PD 125
```

```
Qy          59 DVGACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIA--IPLDNRYERPA--------- 107
               ::   :| |:   ::       |::        |  |:  |||   :  ||
Db         126 ELWEFIADSLKAFID-------EQF--------PQGISEPIPLGFTFSFPASQNKINEGI 170

Qy         108 --------------------------------EVGA------DRLVAAYAARRLYPGPR 128
                                               || |       |||:|     |  |
Db         171 LQRWTKGFDIPNIENHDVVPMLQKQITKRNIHIEVVALINDTTGTLVASY-----YTDPE 225

Qy         129 SLVSVDFGTATTFDCVEGGAY---------LGGLICPGVLSSA--------GALSSRTAK 171
               : : | |||         |||         | | :   :  ||        |:  :
Db         226 TKMGVIFGTGV------NGAYYDVCSDIEKLQGKLSDDIPPSAPMAINCEYGSFDNEHVV 279

Qy         172 LPR--ISLEVEEDSPVIGRSTTTSLNHGFIFGF 202
               |||    : ::|:||  |: |   :: |:  ||
Db         280 LPRTKYDITIDEESPRPGQQTFEKMSSGYYLGF 312


RESULT 2
US-09-360-197-14
; Sequence 14, Application US/09360197
  Patent No. 6287859
  GENERAL INFORMATION:
   APPLICANT:  Bassilana, Frederic
;  APPLICANT:  Lazdunski, Michel
;  APPLICANT:  Waldmann, Rainer
;  APPLICANT:  Deweille, Jan R.
;  TITLE OF INVENTION:  Human and Rat Families of Neuronal Acid-Sensitive
;  TITLE OF INVENTION:  Cationic Channels, Their Cloning and Applications
;  FILE REFERENCE:  989.6706P
;  CURRENT APPLICATION NUMBER: US/09/360,197
;  CURRENT FILING DATE:  1997-07-23
;  PRIOR APPLICATION NUMBER:  09/129,758
;  PRIOR FILING DATE:  1998-08-05
;  PRIOR APPLICATION NUMBER:  60/095,408
;  PRIOR FILING DATE:  1998-08-05
;  NUMBER OF SEQ ID NOS:  22
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 14
;   LENGTH: 531
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-360-197-14

  Query Match             7.4%;  Score 79.5;  DB 4;  Length 531;
  Best Local Similarity   19.9%;  Pred. No. 1;
  Matches   59;  Conservative   29;  Mismatches   94;  Indels  115;  Gaps   11;

Qy           8 FDIGNTNVKIGIAVETAVLTSYVLPTDPG-QTTDSIGLRLLEVLRHAGLGPADVGACVAS 66
               |::|   |:|    |  ::    |   || ||  |   : |  |      |   | | ::
Db         233 FEVG-IRVQIHSQEEPPIIDQLGLGVSPGYQTFVSCQQQQLSFL------PPPWGDCSSA 285

Qy          67 SVVPGVNPL--------------------IRRACE-RYLYRK----LLFAPGDIAIPLDN 101
               |: |   |                    | ||| ||: ||    ::: |||: :
Db         286 SLNPNYEPEPSDPLGSPSPSPSPPYTLMGCRLACETRYVARKCGCRMVYMPGDVPVCSPQ 345

Qy         102 RYERPAEVGAD----------------------------RLVAAYAARRL-----YPG 126
               :|:  |    |                          |  | : ||:|     |
Db         346 QYKNCAHPAIDAMLRKDSCACPNPCASTRYAKELSMVRIPSRAAARFLARKLNRSEAYIA 405

Qy         127 PRSLVSVDFGTATTFDCVEG-------------GAYLGGLICPGVL-------------- 159
                  |    |  |  :: ||              |  :|  |   :|
Db         406 ENVLALDIFFEALNYETVEQKKAYEMSELLGDIGGQMGLFIGASLLTILEILDYLCEVFR 465

Qy         160 ----------------SSAG----ALSSRTAKLPRISLEVEEDSPVIGRSTTTSLNH 196
                             ||       | |   ::| :||     :|    : | | :|
Db         466 DKVLGYFWNRQHSQRHSSTNLLQEGLGSHRTQVPHLSLGPRPPTPPCAVTKTLSASH 522


RESULT 3
US-09-240-639-10
; Sequence 10, Application US/09240639
; Patent No. 6350447
; GENERAL INFORMATION:
;  APPLICANT: Chadwick, Brian Paul
;  APPLICANT: Frischauf, Anna-Maria
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO CD39-LIKE
;  TITLE OF INVENTION: POLYPEPTIDES AND NUCLEIC ACIDS
;  FILE REFERENCE: 9598-066
;  CURRENT APPLICATION NUMBER: US/09/240,639
;  CURRENT FILING DATE: 1998-01-29
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
;   LENGTH: 455
;   TYPE: PRT
;   ORGANISM: P. sativum
US-09-240-639-10

  Query Match             7.3%;  Score 78.5;  DB 4;  Length 455;
  Best Local Similarity   22.0%;  Pred. No. 1;
  Matches   54;  Conservative   37;  Mismatches   62;  Indels   93;  Gaps   14;

Qy           5 FLLFD---------IGN---TNVKIGIAVETAVLTSYVLPTDPGQTTDSIGL----RLLE 48
                 ||||         :||   |: || :  |   ::|| :  | | |    | :    : |:
Db          10 FLLFSMPAITSSQYLGNNLLTSRKIFLKQEE--ISSYAVVFDAGSTGSRIHVYHFNQNLD 67

Qy          49 VLRHAGLGPADVGACVASSVVPGVN-------------PLIRRACERYLYRKLLFAPGD 94
               :| | | |         : : ||::               ||: :|              :
Db          68 LL-HIGKGVE-----YYNKITPGLSSYANNPEQAAKSLIPLLEQA-------------E 107

Qy          95 IAIPLDNRYERPAEVGAD---RLVAAYAARRLYPGPRSLVS------VDFGTATTFDCVE 145
                 :| | : : |  :||    ||:   |: ::    | ::|      |     :  |  :
Db         108 DVVPDDLQPKTPVRLGATAGLRLLNGDASEKILQSVRDMLSNRSTFNVQPDAVSIIDGTQ 167

Qy         146 GGAY-----------------------LGGLICPGVLSSAGALSSRTAK-LPRISLEVE 180
                 |:|                     |||    | :  | |:| :|||  |:::   :
Db         168 EGSYLWVTVNYALGNLGKKYTKTVGVIDLGG----GSVQMAYAVSKKTAKNAPKVA---D 220

Qy         181 EDSPVI 186
                 | | |
Db         221 GDDPYI 226


RESULT 4
US-08-342-930-2
; Sequence 2, Application US/08342930
; Patent No. 5821084
;  GENERAL INFORMATION:
;    APPLICANT:  OLMSTED, ELIZABETH A.
;    APPLICANT:  MAURO, LAURA J.
;    APPLICANT:  DAVIS, ALAN R.
;    APPLICANT:  DIXON, JACK E.
;    TITLE OF INVENTION:  OSTEOBLAST-TESTICULAR PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/342,930
;      FILING DATE:  21-NOV-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20344-20975.00
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1711 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-342-930-2

  Query Match             7.2%;  Score 78;  DB 2;  Length 1711;
  Best Local Similarity   23.2%;  Pred. No. 8.3;
  Matches   62;  Conservative   25;  Mismatches   80;  Indels  100;  Gaps   15;

Qy          4 HFLLFD-IGNTN----VKIGIAVETAV------------------------LTSYVLPT 33
              | :| | :| ||    :: |::  | :                          | :  |:
Db        248 HLMLTDLLGGTNTTAVIRQGVSTHTFLHLSPGTPHELKICASAGPHQIWGPSATEWTYPS 307

Qy         34 DPGQTT------------DSIGLRLLEVLRHAG-------LGPADVGACVASSVVPGVN 73
                |                :| |   ||: :|       ||| :  |     : ||
Db        308 YPSDLVLTPLRNELWASWKAGLGARDGYVLKLSGPMESTSTLGPEECNAVFPGPLPPG-- 365

Qy         74 PLIRRACERYLYRKLLFAPGDIAIP----LDNRYERPAEVGADRL------VAAYAARR- 122
                    |  |:|  | |  :     |      | ||   ||       :    ||
Db        366 -------HYTLQLKVLAGPYDAWVEGSTWLAESAALPREVPGARLWLDGLEASKQPGRRA 418

Qy        123 -LY----PGPRSLVSVDFGTATTFDCVEGGAYLGGLICPG------VLSSAG----ALSS 167
               ||    ||    :||  |      |        ||: ||      : || |    ::|
Db        419 LLYSDDAPGSLGNISVPSGATHVIFC--------GLV-PGAHYRVDIASSTGDISQSISG 469

Qy        168 RTAKLPRISLEVEEDSPVIGRSTTTSL 194
              |: ||  |||      || ||: : |
Db        470 YTSPLPPQSLE------VISRSSPSDL 490


RESULT 5
US-09-129-112-15
; Sequence 15, Application US/09129112
; Patent No. 6465716
; GENERAL INFORMATION:
;  APPLICANT: Etzler, Marilynn E.
;  APPLICANT:  Murphy, Judith B.
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: A No. 6465716 Factor Binding Protein From Legume Roots
;  FILE REFERENCE: 023070-079810US
;  CURRENT APPLICATION NUMBER: US/09/129,112
;  CURRENT FILING DATE:  1998-08-04
;  PRIOR APPLICATION NUMBER: US 08/907,226
;  PRIOR FILING DATE: 1997-08-06
;  NUMBER OF SEQ ID NOS: 19
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
;    LENGTH: 462
;    TYPE: PRT
;    ORGANISM: Lotus japonicus
US-09-129-112-15

  Query Match             7.0%;  Score 75.5;  DB 4;  Length 462;
  Best Local Similarity   23.4%;  Pred. No. 2.4;
  Matches   54;  Conservative   31;  Mismatches   91;  Indels   55;  Gaps   12;

Qy          1 MTQHFLLFD-------IGNT---NVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVL 50
              ||  |:|          :||    | || :       :||| :  | | |   : :   :
Db         14 MTFVFMLMPAISSSQYLGNNILMNRKILLPKNQEPVTSYAVIFDAGSTGSRVHVYNFD-- 71

Qy         51 RHAGLGPADVGACVASSVVPGV-----NPLIRRACERYLYRKLLFAPGDIAIPLDNRYER 105
              ::  | | :        || ||:     ||    | |  :    |    :  :|:  :
Db         72 QNLDLLPVENELEFYDSVKPGLSSYAANP--EEAAESLI---PLLKEAENVVPVSQQPNT 126

Qy        106 PAEVGAD---RLVAAYAARRLYPGPRSLVS------VDFGTATTFDCVEGGAYLGGLI-- 154
              | ::||    ||:   ||  :    | ::|      |     :  |  : |:||   |
Db        127 PVKLGATAGLRLLEGNAAENILQAVRDMLSNRSALNVQSDAVSILDGTQEGSYLWVTINY 186

Qy        155 -----------CPGVLSSAG-------ALSSRTAK-LPRISLEVEEDSPVI 186
                        ||:   |       |:|  |||  |::    | : | |
Db        187 LLGKLGKRFTKTVGVVDLGGGSVQMTYAVSRNTAKNAPKVP---EGEDPYI 234


RESULT 6
US-08-276-919-4
; Sequence 4, Application US/08276919
; Patent No. 5589579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RICHARDS, MEDLOCK & ANDREWS
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,919
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelligan, Mark C.
;      REGISTRATION NUMBER:  36,389
;      REFERENCE/DOCKET NUMBER:  B35792
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-939-4500
;      TELEFAX:  214-939-4600
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-919-4

  Query Match             7.0%;  Score 75;  DB 1;  Length 325;
  Best Local Similarity   24.9%;  Pred. No. 1.7;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||              | |  ::   | : |     :  | || :| ||   |   |:
Db        109 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 154

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                       |      : ||| |    |  | | | ||  |  |      |   :  |
Db        155 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 214

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD-------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||             |   |
Db        215 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 271

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        272 SLGIILS---LALAGIL 285


RESULT 7
US-08-276-919-13
; Sequence 13, Application US/08276919
; Patent No. 5589579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
```

```
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RICHARDS, MEDLOCK & ANDREWS
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,919
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelligan, Mark C.
;      REGISTRATION NUMBER:  36,389
;      REFERENCE/DOCKET NUMBER:  B35792
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-939-4500
;      TELEFAX:  214-939-4600
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-919-13

  Query Match             7.0%;  Score 75;  DB 1;  Length 325;
  Best Local Similarity   24.9%;  Pred. No. 1.7;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||              | |  ::   | : |     :  | || :| ||   |   |:
Db        109 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 154

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |  :  |
Db        155 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 214

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||            |   |
Db        215 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 271

          149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        272 SLGIILS---LALAGIL 285


RESULT 8
US-08-776-088-4
; Sequence 4, Application US/08776088
; Patent No. 5773579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SIDLEY & AUSTIN
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/776,088
;      FILING DATE:  19-JUL-95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eugenia S. Hansen
;      REGISTRATION NUMBER:  31,966
;      REFERENCE/DOCKET NUMBER:  10365/05011
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-981-3300
;      TELEFAX:  214-981-3400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-776-088-4

  Query Match             7.0%;  Score 75;  DB 1;  Length 325;
  Best Local Similarity   24.9%;  Pred. No. 1.7;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||              | |  ::   | : |     :  | || :| ||   |   |:
Db        109 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 154

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |  :  |
Db        155 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 214

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||            |   |
Db        215 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 271

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        272 SLGIILS---LALAGIL 285


RESULT 9
US-08-776-088-18
; Sequence 18, Application US/08776088
; Patent No. 5773579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SIDLEY & AUSTIN
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/776,088
;      FILING DATE:  19-JUL-95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eugenia S. Hansen
;      REGISTRATION NUMBER:  31,966
;      REFERENCE/DOCKET NUMBER:  10365/05011
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  214-981-3300
;      TELEFAX:  214-981-3400
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-776-088-18

  Query Match             7.0%;  Score 75;  DB 1;  Length 325;
  Best Local Similarity   24.9%;  Pred. No. 1.7;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        109 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 154

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |   :  |
Db        155 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 214

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::. ::|:|   |  ||      | || ::: :|     ||            |   |
Db        215 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 271

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        272 SLGIILS---LALAGIL 285


RESULT 10
PCT-US95-09145A-4
; Sequence 4, Application PC/TUS9509145A
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RICHARDS, MEDLOCK & ANDREWS
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09145A
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John A. Harre
;      REGISTRATION NUMBER:  37,345
;      REFERENCE/DOCKET NUMBER:  B35792CIPPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-939-4500
;      TELEFAX:  214-939-4600
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-09145A-4

  Query Match             7.0%;  Score 75;  DB 5;  Length 325;
  Best Local Similarity   24.9%;  Pred. No. 1.7;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        109 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 154

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |   :  |
Db        155 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 214

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||            |   |
Db        215 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 271

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        272 SLGIILS---LALAGIL 285


RESULT 11
PCT-US95-09145A-18
; Sequence 18, Application PC/TUS9509145A
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RICHARDS, MEDLOCK & ANDREWS
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/09145A
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  John A. Harre
;      REGISTRATION NUMBER:  37,345
;      REFERENCE/DOCKET NUMBER:  B35792CIPPCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-939-4500
;      TELEFAX:  214-939-4600
;  INFORMATION FOR SEQ ID NO:  18:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-09145A-18

  Query Match             7.0%;  Score 75;  DB 5;  Length 325;
  Best Local Similarity   24.9%;  Pred. No. 1.7;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        109 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 154

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |   :  |
Db        155 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 214

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||            |   |
Db        215 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 271

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
```

```
Db        272 SLGIILS---LALAGIL 285


RESULT 12
US-08-276-919-2
; Sequence 2, Application US/08276919
; Patent No. 5589579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  16
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  RICHARDS, MEDLOCK & ANDREWS
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/276,919
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Nelligan, Mark C.
;      REGISTRATION NUMBER:  36,389
;      REFERENCE/DOCKET NUMBER:  B35792
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-939-4500
;      TELEFAX:  214-939-4600
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  354 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-276-919-2

  Query Match             7.0%;  Score 75;  DB 1;  Length 354;
  Best Local Similarity   24.9%;  Pred. No. 1.9;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        138 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 183

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                       |     : ||| |    |  |  | | ||  |  |       |  :  |
Db        184 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 243

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||             |   |
Db        244 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 300

Qy        149 YLGGLICPGVLSSAGAL 165
                || ::    |: || |
Db        301 SLGIILS---LALAGIL 314


RESULT 13
US-08-776-088-2
; Sequence 2, Application US/08776088
; Patent No. 5773579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SIDLEY & AUSTIN
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/776,088
;      FILING DATE:  19-JUL-95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eugenia S. Hansen
;      REGISTRATION NUMBER:  31,966
;      REFERENCE/DOCKET NUMBER:  10365/05011
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-981-3300
;      TELEFAX:  214-981-3400
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  354 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-776-088-2

  Query Match             7.0%;  Score 75;  DB 1;  Length 354;
  Best Local Similarity   24.9%;  Pred. No. 1.9;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        138 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 183

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                       |     : ||| |    |  |  | | ||  |  |       |  :  |
Db        184 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 243

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||             |   |
Db        244 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 300

Qy        149 YLGGLICPGVLSSAGAL 165
                || ::    |: || |
Db        301 SLGIILS---LALAGIL 314


RESULT 14
US-08-776-088-6
; Sequence 6, Application US/08776088
; Patent No. 5773579
;  GENERAL INFORMATION:
;    APPLICANT:  Torczynski, Richard M.
;    APPLICANT:  Bollon, Arthur P.
;    TITLE OF INVENTION:  Lung Cancer Marker
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SIDLEY & AUSTIN
;      STREET:  1201 Elm Street, Suite 4500
;      CITY:  Dallas
;      STATE:  TX
;      COUNTRY:  US
;      ZIP:  75270-2197
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/776,088
;      FILING DATE:  19-JUL-95
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Eugenia S. Hansen
;      REGISTRATION NUMBER:  31,966
;      REFERENCE/DOCKET NUMBER:  10365/05011
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  214-981-3300
;      TELEFAX:  214-981-3400
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  354 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-776-088-6

  Query Match             7.0%;  Score 75;  DB 1;  Length 354;
  Best Local Similarity   24.9%;  Pred. No. 1.9;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        138 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 183

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |   :  |
Db        184 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 243

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||             |   |
Db        244 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 300

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        301 SLGIILS---LALAGIL 314


RESULT 15
US-09-325-320-2
; Sequence 2, Application US/09325320
; Patent No. 6117981
; GENERAL INFORMATION:
;  APPLICANT: Torczynski, Richard M.
;  APPLICANT: Bollon, Arthur P.
   TITLE OF INVENTION: HYBRIDOMAS FOR LUNG CANCER MARKER AND MONOCLONAL
;  TITLE OF INVENTION: ANTIBODIES THEREOF
;  FILE REFERENCE: 10365/06101
;  CURRENT APPLICATION NUMBER: US/09/325,320
;  CURRENT FILING DATE: 1999-06-03
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 354
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-325-320-2

  Query Match             7.0%;  Score 75;  DB 3;  Length 354;
  Best Local Similarity   24.9%;  Pred. No. 1.9;
  Matches   49;  Conservative   20;  Mismatches   74;  Indels   54;  Gaps   10;

Qy          3 QHFLLFDIGNTNVKIGIAVETAVL--TSYVLPTDPGQTTDSIGLRLLEVLRHAG-LGPA- 58
              |||               | |  ::   | : |     :  | || :| ||   |   |:
Db        138 QHF--------------AAELHIVHYNSDLYPDASTASNKSEGLAVLAVLIEMGSFNPSY 183

Qy         59 --------DVGACVASSVVPGVN---PLIRRACERYLYRKLLFAPGDIAIPLDNRYERPA 107
                    |      : ||| |    |  |  | | ||  |  |      |   :  |
Db        184 DKIFSHLQHVKYKGQEAFVPGFNIEELLPERTAEYYRYRGSLTTPPCNPTVLWTVFRNPV 243

Qy        108 EVGADRLVAAYAARRLY------PGPRSLVSVDFGTATTFD------------CVEGGA 148
              ::  ::|:|   |  ||      | || ::: :|     ||             |   |
Db        244 QISQEQLLALETA--LYCTHMDDPSPREMIN-NFRQVQKFDERLVYTSFSQVQVCTAAGL 300

Qy        149 YLGGLICPGVLSSAGAL 165
               || ::    |: || |
Db        301 SLGIILS---LALAGIL 314


Search completed: June 24, 2003, 22:10:48
Job time : 9.60143 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 11.7818 Seconds
(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453A-59
Perfect score: 1077
Sequence: 1 MTQHFLLFDIGNTNVKIGIA.....SLNHGFIFGFAAMTEGVLAA 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

earched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1077	100.0	212	9	US-09-813-453A-59	Sequence 59, Appl
2	379.5	35.2	262	9	US-09-813-453A-45	Sequence 45, Appl
3	352.5	32.7	258	9	US-09-813-453A-49	Sequence 49, Appl
4	348.5	32.4	219	9	US-09-813-453A-57	Sequence 57, Appl
5	338.5	31.4	233	9	US-09-813-453A-17	Sequence 17, Appl
6	338.5	31.4	255	9	US-09-813-453A-7	Sequence 7, Appli
7	338.5	31.4	258	9	US-09-813-453A-2	Sequence 2, Appli
8	330.5	30.7	254	9	US-09-813-453A-47	Sequence 47, Appl
9	327.5	30.4	250	9	US-09-813-453A-3	Sequence 3, Appli
10	315.5	29.3	256	9	US-09-813-453A-55	Sequence 55, Appl
11	307.5	28.6	260	9	US-09-813-453A-51	Sequence 51, Appl
12	280	26.0	265	9	US-09-813-453A-4	Sequence 4, Appli
13	271	25.2	262	9	US-09-813-453A-8	Sequence 8, Appli
14	240.5	22.3	258	9	US-09-813-453A-6	Sequence 6, Appli
15	221	20.5	246	9	US-09-813-453A-9	Sequence 9, Appli
16	214	19.9	257	9	US-09-813-453A-53	Sequence 53, Appl
17	209.5	19.5	272	9	US-09-712-363-276	Sequence 276, App
18	209.5	19.5	272	9	US-09-813-453A-5	Sequence 5, Appli
19	192.5	17.9	273	9	US-09-813-453A-10	Sequence 10, Appl
20	180.5	16.8	241	9	US-09-813-453A-63	Sequence 63, Appl
21	174	16.2	244	9	US-09-813-453A-41	Sequence 41, Appl
22	138	12.8	267	9	US-09-813-453A-15	Sequence 15, Appl
23	136.5	12.7	592	9	US-09-813-453A-43	Sequence 43, Appl
24	134	12.4	249	9	US-09-813-453A-70	Sequence 70, Appl
25	133.5	12.4	460	9	US-09-813-453A-39	Sequence 39, Appl
26	132.5	12.3	592	9	US-09-813-453A-22	Sequence 22, Appl
27	120.5	11.2	229	9	US-09-813-453A-12	Sequence 12, Appl
28	114.5	10.6	249	9	US-09-813-453A-61	Sequence 61, Appl
29	111	10.3	262	9	US-09-813-453A-11	Sequence 11, Appl
30	107	9.9	248	9	US-09-813-453A-20	Sequence 20, Appl
31	104	9.7	242	9	US-09-813-453A-65	Sequence 65, Appl
32	102	9.5	223	9	US-09-895-913A-74	Sequence 74, Appl
33	102	9.5	223	9	US-09-813-453A-14	Sequence 14, Appl
34	102	9.5	223	9	US-09-813-453A-67	Sequence 67, Appl
35	92.5	8.6	257	9	US-09-813-453A-13	Sequence 13, Appl
36	88	8.2	275	9	US-10-106-698-6389	Sequence 6389, Ap
37	85	7.9	209	9	US-09-813-453A-21	Sequence 21, Appl
38	82.5	7.7	486	10	US-09-801-368-154	Sequence 154, App
39	82	7.6	250	9	US-10-167-015-2	Sequence 2, Appli
40	81	7.5	376	9	US-09-738-626-6189	Sequence 6189, Ap
41	80.5	7.5	2201	12	US-10-029-907-3	Sequence 3, Appli
42	79.5	7.4	674	9	US-10-086-464-14	Sequence 14, Appl
43	79	7.3	324	10	US-09-815-242-5117	Sequence 5117, Ap
44	77	7.1	4999	9	US-09-976-059-15	Sequence 15, Appl
45	76	7.1	731	9	US-10-086-464-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
US-09-813-453A-59
; Sequence 59, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 59
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Desulfovibrio vulgaris
US-09-813-453A-59

Query Match 100.0%; Score 1077; DB 9; Length 212;
Best Local Similarity 100.0%; Pred. No. 1e-102;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60

Qy         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120

Qy        121 RRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVE 180

Qy        181 EDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212
              ||||||||||||||||||||||||||||||||
```

```
Db        181 EDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212


RESULT 2
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
  SEQ ID NO 45
    LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
US-09-813-453A-45

  Query Match           35.2%;  Score 379.5;  DB 9;  Length 262;
  Best Local Similarity  40.8%;  Pred. No. 5.4e-31;
  Matches   84;  Conservative   41;  Mismatches   74;  Indels    7;  Gaps    5;

Qy          7 LFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVAS 66
              : |:|||| :|:  |  :   : : ||  :| |  |: : ::| | ||   ||   : |
Db          4 VLDVGNTNAVLGVFEEGELRQHWRMETDRHKTEDEYGMLVKQLLEHEGLSFEDVKGIIVS 63

Qy         67 SVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||| :   : | ||:|   : |:  || |   |: :|| | ||||||:| | |   ||
Db         64 SVVPPIMFALERMCEKYFKIKPLVVGPG-IKTGLNIKYENPREVGADRIVNAVAGIHLYG 122

Qy        126 GPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
              |  |: |||||||:  : |   |:||:| ||:: || || || ||||||  |: : |
Db        123 SP--LIIVDFGTATTYCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRI--EITKPSS 178

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:|::| :::  | ::|:    ||::
Db        179 VVGKNTVSAMQSGILYGYVGQVEGIV 204


RESULT 3
S-09-813-453A-49
  Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 258
;   TYPE: PRT
;   ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

  Query Match           32.7%;  Score 352.5;  DB 9;  Length 258;
  Best Local Similarity  37.7%;  Pred. No. 3.1e-28;
  Matches   78;  Conservative   42;  Mismatches   80;  Indels    7;  Gaps    5;

Qy          7 LFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVAS 66
              : |:|||| :|:     :   : : |   :| |  |: :  :| | ||  :|:    : |
Db          4 VLDVGNTNTVLGVYDGDELKHHWRIETSRSKTEDEYGMMIKALLNHVGLQFSDIRGIIIS 63

Qy         67 SVVPGVNPLIRRACERYLYRK-LLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||| :   : | | :| : | |:  || |   || :|: | |||||||:| | |   ||
Db         64 SVVPPIMFALERMCLKYFHIKPLIVGPG-IKTGLDIKYDNPREVGADRIVNAVAGIHLYG 122

Qy        126 GPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
              |  |: |||||||:  : |   |:|| | ||:: |  || :| ||||||  |:
Db        123 SP--LIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRI--EIARPDD 178

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              :||:: :::  | ::|:    ||:::
Db        179 IIGKNTVSAMQAGILYGYVGQVEGIVS 205


RESULT 4
US-09-813-453A-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
;   LENGTH: 219
;   TYPE: PRT
;   ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

  Query Match           32.4%;  Score 348.5;  DB 9;  Length 219;
  Best Local Similarity  38.5%;  Pred. No. 6.4e-28;
  Matches   80;  Conservative   39;  Mismatches   84;  Indels    5;  Gaps    3;

Qy          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60
              |::  :  |||||:| |||     :| :: | :   :  |     || :|:|||: | ::
Db          1 MSEKLVAVDIGNTSVNIGIFEGEKLLANWHLGSVAQRMADEYASLLLGLLQHAGIHPEEL 60

Qy         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120
                 :  |||| :       : |     |     |   :  | : | |||||:| | ||
Db         61 NRVIMCSVVPPLTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAAAA 120

Qy        121 RRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEV 179
              | ||||  : : || ||||||| : |||||:|| | ||: :|| |:: :|:|||:|  |:
Db        121 RVLYPG--ACIIVDMGTATTFDTLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKI--EI 176

Qy        180 EEDSPVIGRSTTTSLNHGFIFGFAAMTE 207
              : ||| :| :::  |  ||:  : |
Db        177 IRPAKVIGSNTVSAMQSGIYFGYIGLVE 204


RESULT 5
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
```

; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-17

Query Match 31.4%; Score 338.5; DB 9; Length 233;
Best Local Similarity 37.2%; Pred. No. 7.4e-27;
Matches 77; Conservative 40; Mismatches 83; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy 66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
Db 178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204

RESULT 6
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

Query Match 31.4%; Score 338.5; DB 9; Length 255;
Best Local Similarity 38.6%; Pred. No. 8.4e-27;
Matches 80; Conservative 38; Mismatches 82; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGILINELFRLAGLGLDQIRAVII 62

Qy 66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPLTGVLERLSLGYFGMRPLVVGPG-IKTGMPIQYDNPREVGADRIVNAVAGYEKY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 --RTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPRV--DIIRPS 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
Db 178 AIIARNTVNSMQAGIYYGYVGLVDEIV 204

RESULT 7
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 31.4%; Score 338.5; DB 9; Length 258;
Best Local Similarity 37.2%; Pred. No. 8.5e-27;
Matches 77; Conservative 40; Mismatches 83; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy 66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
Db 178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204

RESULT 8
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47

```
;   LENGTH: 254
;   TYPE: PRT
;   ORGANISM: Bacillus halodurans
US-09-813-453A-47

  Query Match             30.7%;  Score 330.5;  DB 9;  Length 254;
  Best Local Similarity   36.7%;  Pred. No. 5.5e-26;
  Matches   76;  Conservative   42;  Mismatches   82;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|:  :  ::  : | |   :| |   : :  :  ||||   |:  |
Db          3 LVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGIVI 62

Qy         66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   : : |::| :   ::  || |   |: :|: | ||||||:| | ||  ||
Db         63 SSVVPPMMFSLEQMCKKYFHVTPMIIGPG-IKTGLNIKYDNPKEVGADRIVNAVAAIELY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |   : ||||||||:  : |   | ||:| ||:: |  ||  | :|||||  |: :
Db        122 GYP--AIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRI--EIAKPK 177

          184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:| :|  |:  |  :|: :  :||:
Db        178 QVVGTNTIDSMQSGIFYGYVSQVDGVV 204


RESULT 9
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

  Query Match             30.4%;  Score 327.5;  DB 9;  Length 250;
  Best Local Similarity   37.7%;  Pred. No. 1.1e-25;
  Matches   75;  Conservative   42;  Mismatches   75;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :||  :| :   : | ||  :: |  |:::: : :   | |  |   :
Db         20 LVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGVII 79

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :      :|     |:  || |   :: :|: | |||||||:| | ||  :|
Db         80 SSVVPNIMYSLEHMIRKYFKINPLVVGPG-IKTGINIKYDNPKEVGADRIVNAVAAHEIY 138

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |||: :||||||||  | | | |||| ||||:  |: ||  : |||||:  |: : :
Db        139 --KRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRV--ELIKPA 194

Qy        184 PVIGRSTTTSLNHGFIFGF 202
              | ::| :|:  | :: :
Db        195 YAICKNTISSIQSGIVYRY 213


RESULT 10
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
;   LENGTH: 256
;   TYPE: PRT
;   ORGANISM: Clostridium difficile
US-09-813-453A-55

  Query Match             29.3%;  Score 315.5;  DB 9;  Length 256;
  Best Local Similarity   35.1%;  Pred. No. 1.9e-24;
  Matches   73;  Conservative   46;  Mismatches   82;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |:||:|||: :||     ::  : : ||  :|:|  |: :  :  :  : :|:   :
Db          3 LVFDVGNTNMVLGIYKGDKLVNYWRIKTDREKTSDEYGILISNLFDYDNVNISDIDDVII 62

Qy         66 SSVVPGVNPLIRRACERYLYRK-LLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| |   :   | :|  :: |:  || |   |: :|: | :|||||:| | |    |
Db         63 SSVVPNVMHSLENFCIKYCKKQPLIVGPG-IKTGLNIKYDNPKQVGADRIVNAVAGIEKY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  |  |||||||||  : | | |||| | ||:  |: ||    :||||:  |: :
Db        122 GAPSIL--VDFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPRV--ELAKPG 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              | :|| :::  | |:|:  : : :::
Db        178 MTICKSTVSAMQSGIIYGYVGLVDKIIS 205


RESULT 11
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-813-453A-51

  Query Match             28.6%;  Score 307.5;  DB 9;  Length 260;
  Best Local Similarity   36.2%;  Pred. No. 1.3e-23;
  Matches   75;  Conservative   38;  Mismatches   91;  Indels    3;  Gaps    2;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |  : ||||    |   : :  :   |:  :| |   : | ::|   |||   : | :
```

```
Db          3 LAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAVII 62

Qy         66 SSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125

Db         63 SSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYP 122

Qy        126 GPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184

Db        123 GP--LVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211

Db        181 IVGTDTVSAMQSGVFWGYISLIEGLVA 207


RESULT 12
US-09-813-453A-4
; Sequence 4, Application US/09813453A
  Patent No. US20020168681A1
  GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

  Query Match           26.0%;  Score 280;  DB 9;  Length 265;
  Best Local Similarity  35.5%;  Pred. No. 8.9e-21;
  Matches   76;  Conservative   36;  Mismatches   86;  Indels   16;  Gaps    8;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSI-----GLRLLEVLRHAGLGPADV 60

Db          3 LTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDGID 62

Qy         61 GACVASSVVPGVNPLIRRACERYL--YRKLLFAPG-DIAIPLDNRYERPAEVGADRLVAA 117

Db         63 GIAICAT-VPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPI--LTDHPKEVGADRIINA 119

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRIS 176

Db        120 VAAVELYGGP--AIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKI- 176

Qy        177 LEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210

Db        177 -EVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVV 209


RESULT 13
US-09-813-453A-8
; Sequence 8, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Deinococcus radiopugnans
US-09-813-453A-8

  Query Match           25.2%;  Score 271;  DB 9;  Length 262;
  Best Local Similarity  36.3%;  Pred. No. 7.4e-20;
  Matches   77;  Conservative   32;  Mismatches   87;  Indels   16;  Gaps    7;

Qy          6 LLFDIGNTNVKIGIAVETAVLT-SYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV-GAC 63

Db          7 LAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAG---APIPRAA 63

Qy         64 VASSVVPGVNPLIRRACERYLYRKLLFAPG----DIAIPLDNRYERPAEVGADRLVAAYA 119

Db         64 VLSSVAPPVGENYALALKRHFMIDAFAVSAENLPDVTVELDT----PGSVGADRLCNLFG 119

Qy        120 ARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLE 178

Db        120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy        179 VEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210

Db        180 APETA--IGKNTVHALQSGLVFGYAEMVDGLL 209


RESULT 14
US-09-813-453A-6
; Sequence 6, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 258
;   TYPE: PRT
;   ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

  Query Match           22.3%;  Score 240.5;  DB 9;  Length 258;
  Best Local Similarity  33.3%;  Pred. No. 9.8e-17;
  Matches   69;  Conservative   30;  Mismatches  101;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65

Db          3 LCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLNTLMQLKGL-QGRISEAII 61

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124

Db         62 SSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYDRH 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183

Db        122 GG--DLIVVDFGTATTFDVVAPDGAYIGGVIAPGVNLSLEALHMAAAALPHV--DVTKPQ 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
```

```
              ||| :|   :  |  :|:  : ||::
Db        178 GVIGTNTVACIQSGVYWGYIGLVEGIV 204


RESULT 15
US-09-813-453A-9
; Sequence 9, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
   NUMBER OF SEQ ID NOS: 77
   SOFTWARE: PatentIn Ver. 2.0
  SEQ ID NO 9
;   LENGTH: 246
;   TYPE: PRT
;   ORGANISM: Thermotoga maritima
US-09-813-453A-9

  Query Match           20.5%;  Score 221;  DB 9;  Length 246;
  Best Local Similarity   31.9%;  Pred. No. 9.2e-15;
  Matches   68;  Conservative   31;  Mismatches   92;  Indels   22;  Gaps    7;

Qy          5 FLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPA--DVGA 62
              :|| |:|||:    |  :      : | |   || | :   |     |  || |  ::
Db          2 YLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHL-----HPLLGDAMREIKG 56

Qy         63 CVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDN-----RYERPAEVGADRLVAA 117
               :|||| | :| |  ::|      |    | :   |       : |:||||||:
Db         57 IGVASVVPTQNTVIERFSQKY------FHISPIWVKAKNGCVKWNVKNPSEVGADRVANV 110

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISL 177
              |  : | |   :: :| ||||| | |  |:| || | ||      :|   ||||| :
Db        111 VAFVKEY-GKNGII-IDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLV-- 166

Qy        178 EVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              ||:    |:|: |  ::  | : |     ||::
Db        167 EVKPADFVVGKDTEENIRLGVVNGSVYALEGII 199


arch completed: June 24, 2003, 22:29:06
b time : 11.7818 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 18.6485 Seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453A-59
Perfect score: 1077
Sequence: 1 MTQHFLLFDIGNTNVKIGIA.....SLNHGFIFGFAAMTEGVLAA 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1077	100.0	212	2	O32514	O32514 desulfovibr
2	382.5	35.5	259	16	Q92F54	Q92f54 listeria in
3	380.5	35.3	259	16	Q8YAC5	Q8yac5 listeria mo
4	353.5	32.8	259	16	Q8XHL5	Q8xhl5 clostridium
5	352.5	32.7	258	2	Q9F985	Q9f985 bacillus st
6	351.5	32.6	255	16	Q8R7M2	Q8r7m2 thermoanaer
7	335.5	31.2	273	16	Q97EB4	Q97eb4 clostridium
8	330.5	30.7	254	16	Q9KGH5	Q9kgh5 bacillus ha
9	307.5	28.6	261	16	Q9A6Z1	Q9a6z1 caulobacter
10	280	26.0	265	16	Q9X8N6	Q9x8n6 streptomyce
11	271	25.2	262	16	Q9RX54	Q9rx54 deinococcus
12	221	20.5	246	16	Q9WZY5	Q9wzy5 thermotoga
13	219.5	20.4	256	16	Q8RFE4	Q8rfe4 fusobacteri
14	210.5	19.5	274	16	Q9CD56	Q9cd56 mycobacteri
15	209.5	19.5	272	16	O06282	O06282 mycobacteri
16	192.5	17.9	273	16	O83446	O83446 treponema p
17	186	17.3	295	16	Q8Y2M4	Q8y2m4 ralstonia s
18	145.5	13.5	276	16	Q8YQD7	Q8yqd7 anabaena sp
19	136.5	12.7	592	16	Q9JXF1	Q9jxf1 neisseria m
20	132.5	12.3	592	16	Q9JWI7	Q9jwi7 neisseria m
21	120.5	11.2	229	16	O67753	O67753 aquifex aeo
22	111	10.3	262	16	O51477	O51477 borrelia bu
23	107	9.9	248	16	Q9HWC1	Q9hwc1 pseudomonas
24	104	9.7	242	16	Q9PCI4	Q9pci4 xylella fas
25	102	9.5	223	16	O25533	O25533 helicobacte
26	97.5	9.1	699	10	Q9XI96	Q9xi96 arabidopsis
27	97	9.0	223	16	Q9ZKY6	Q9zky6 helicobacte
28	97	9.0	400	16	Q9RRS2	Q9rrs2 deinococcus
29	96.5	9.0	308	17	Q9YA95	Q9ya95 aeropyrum p
30	94.5	8.8	577	17	Q9YEY4	Q9yey4 aeropyrum p
31	94	8.7	345	2	Q8RTT9	Q8rtt9 uncultured
32	93.5	8.7	871	16	Q9L2D2	Q9l2d2 streptomyce
33	93	8.6	441	2	Q9LBD8	Q9lbd8 rhodobacter
34	92.5	8.6	257	16	P74045	P74045 synechocyst
35	91.5	8.5	312	2	Q52016	Q52016 pseudomonas
36	91.5	8.5	458	16	Q9KVU6	Q9kvu6 vibrio chol
37	91	8.4	306	2	O51855	O51855 pseudomonas
38	90.5	8.4	252	4	Q8WXQ0	Q8wxq0 homo sapien
39	90.5	8.4	546	2	Q9KVZ8	Q9kvz8 xanthomonas
40	90.5	8.4	571	2	Q8VT54	Q8vt54 xanthomonas
41	89.5	8.3	308	17	Q8TMD8	Q8tmd8 methanosarc
42	89.5	8.3	312	2	O85978	O85978 sphingomona
43	89	8.3	380	16	Q9KPM8	Q9kpm8 vibrio chol
44	89	8.3	387	16	Q9X7Y0	Q9x7y0 streptomyce
45	88.5	8.2	752	10	Q9LQC5	Q9lqc5 arabidopsis

ALIGNMENTS

RESULT 1
O32514
ID O32514 PRELIMINARY; PRT; 212 AA.
AC O32514;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Simillar to Bacillus subtilis.
OS Desulfovibrio vulgaris (strain Miyazaki).
OC Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC Desulfovibrio.
OX NCBI_TaxID=883;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MIYAZAKI;
RA Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,
RA Kohno K., Akutsu H., Kumagai I., Nakaya T.;
RT "Sequence analyses of two ferredoxin genes and their flanking regions
RT from Desulfovibrio vulgaris (Miyazaki F).";
RL J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
DR EMBL; AB005550; BAA21476.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
SQ SEQUENCE 212 AA; 22336 MW; 23E1789973A344D7 CRC64;

Query Match 100.0%; Score 1077; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 1.1e-86;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60

Qy         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120

Qy        121 RRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 RRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVE 180

Qy        181 EDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212
              ||||||||||||||||||||||||||||||||
Db        181 EDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212


RESULT 2
Q92F54
ID   Q92F54      PRELIMINARY;      PRT;   259 AA.
AC   Q92F54;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein lin0253.
GN   LIN0253.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596164; CAC95486.1; -.
DR   ListiList; LIN00253; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28227 MW;  554B03A0C0EFA64F CRC64;

  Query Match             35.5%;  Score 382.5;  DB 16;  Length 259;
  Best Local Similarity   39.9%;  Pred. No. 1.1e-25;
  Matches   83;  Conservative   40;  Mismatches   78;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|:  |  :|  : : ||  :|:| :|: :|    :| | |:|:   :
Db          3 LVIDVGNTNCTVGVYKEQKLLRHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGIII 62

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :   | ||   | |:  || |   |: : : | |:|:||:| | ||   |
Db         63 SSVVPPIMHAMETMCVRYFNIRPLIVGPG-IKTGLNLKVDNPREIGSDRIVNAVAASEEY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  :: |||||||||  : | | | || | ||:: |  || :| |||||:  :: | |
Db        122 GTP--VIVVDFGTATTFCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPRV--DIAESS 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              :||:||  |:  |  :||   ||::|
Db        178 QIIGKSTVASMQAGIFYGFIGQCEGIIA 205


RESULT 3
Q8YAC5
ID   Q8YAC5      PRELIMINARY;      PRT;   259 AA.
AC   Q8YAC5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein lmo0221.
GN   LMO0221.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-E / SEROVAR 1/2A;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591974; CAD00748.1; -.
DR   MEROPS; M41.009; -.
DR   ListiList; LMO00221; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28187 MW;  F106B049D80025B7 CRC64;

  Query Match             35.3%;  Score 380.5;  DB 16;  Length 259;
  Best Local Similarity   39.4%;  Pred. No. 1.6e-25;
  Matches   82;  Conservative   42;  Mismatches   77;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|:  :  :|  : : ||  :|:| :|: :|    :| | |:|:   :
Db          3 LVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGIII 62

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :   | ||   | |:  || |   |: : : | |:|:||:| | ||   |
Db         63 SSVVPPIMHAMETMCVRYFNIRPLIVGPG-IKTGLNLKVDNPREIGSDRIVNAVAASEEY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  :: |||||||||  : | | | || | ||:: |  || :| |||||:  :: | |
Db        122 GTP--VIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRV--DIAESS 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              :||:||  |:  |  :||   ||::|
Db        178 QIIGKSTVSSMQAGIFYGFVGQCEGIIA 205


RESULT 4
Q8XHL5
ID   Q8XHL5      PRELIMINARY;      PRT;   259 AA.
AC   Q8XHL5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CPE2468.
GN   CPE2468.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
```

```
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003194; BAB82174.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28819 MW;  0D5FEA3B7A145E10 CRC64;

  Query Match             32.8%;  Score 353.5;  DB 16;  Length 259;
  Best Local Similarity   39.2%;  Pred. No. 3.8e-23;
  Matches   82;  Conservative   40;  Mismatches   76;  Indels   11;  Gaps    6;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              || |:|||: :||      : |: : ||  :|:|   ::::::   | | | ||    :
Db          3 LLIDVGNTNIVLGIHDNEKYIASWRISTDSKKTSDEYSIQVMQLFNQAKLNPEDVEGIII 62

Qy         66 SSVVPGVNPLIR---RACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARR 122
              ||||| :   :    | |  :    ::  || |   :: :|: | ||||||:| | ||
Db         63 SSVVPNIMHSLENMVRKC--FCKEPIVVGPG-IKTGINIKYDNPKEVGADRIVNAVAAFE 119

Qy        123 LYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEE 181
               :  |  :: :||||||||  : | | |||| ||||:  || ||  | ||||||  |:|:
Db        120 KHKKP--MIIIDFGTATTFCAITEKGDYLGGNICPGIQISADALFERAAKLPRI--ELEK 175

Qy        182 DSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                || ::| ||:  | |:|:    | ::
Db        176 PKSVICKNTVTSMQAGIIYGYIGKVEYIV 204


RESULT 5
Q9F985
ID   Q9F985      PRELIMINARY;      PRT;   258 AA.
AC   Q9F985;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative 32 kDa replication protein.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=V;
RA   Vasquez C., Pichuantes S., Saavedra C.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF198621; AAG28531.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
SQ   SEQUENCE   258 AA;  28101 MW;  507B55D695095855 CRC64;

  Query Match             32.7%;  Score 352.5;  DB 2;  Length 258;
  Best Local Similarity   37.2%;  Pred. No. 4.6e-23;
  Matches   77;  Conservative   44;  Mismatches   79;  Indels    7;  Gaps    5;

Qy          7 LFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVAS 66
              : |:|||| :|:     :  : : |  |:| |  |: :  :| | ||  :|:   : |
Db          4 VLDVGNTNTVLGVYDGDELKHHWRIETSRGKTEDEYGMTIKALLNHVGLQFSDIDGIIIS 63

Qy         67 SVVPGVNPLIRRACERYLYRK-LLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||| :   : | | :| : | ::  || |   |: :|: | ||||||:| | |   ||
Db         64 SVVPPIMFALERMCLKYFHIKPIIVGPG-IKTGLNIKYDNPREVGADRIVNAVAGIHLYG 122

Qy        126 GPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
               |  |: ||||||||:  : |   |:|| | ||:: |  || :| ||||||  |:
Db        123 SP--LIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRI--EIARPDD 178

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              :||::| :::  | ::|:    ||:::
Db        179 IIGKNTVSAMQAGILYGYVGQVEGIVS 205


RESULT 6
Q8R7M2
ID   Q8R7M2      PRELIMINARY;      PRT;   255 AA.
AC   Q8R7M2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcriptional regulator, homologs of Bvg accessory
DE   factor.
GN   TTE2381.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013180; AAM25520.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   255 AA;  27816 MW;  C3C620ECBC8CA6ED CRC64;

  Query Match             32.6%;  Score 351.5;  DB 16;  Length 255;
  Best Local Similarity   38.6%;  Pred. No. 5.5e-23;
  Matches   80;  Conservative   45;  Mismatches   75;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              | ||:|||: :|:     :| |: : ||  :| |  |: : ::: : |:   ::   :
Db          3 LAFDVGNTNIVMGVFKGKKLLHSFRISTDKNKTYDEYGMLVNQLIGYNGISLTEIDDVII 62

Qy         66 SSVVPG-VNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              |||||  :| |   : : :  : ::  || |  :: :|: | ||||||:| | ||  ||
Db         63 SSVVPPLMNTLQVMSLKYFRTKPIVVGPG-IKTGINIKYDNPKEVGADRIVNAVAAYELY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
               ||  :: :||||||||  : | | ||||:| ||:: || ||  ||||||:| |   :
Db        122 GGP--VIVIDFGTATTFCAISEKGEYLGGIIAPGLMISADALFQRTAKLPKIDL--TKPP 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                 || |:|  |:  | |:|   | : ::
Db        178 TVINRNTVASMQSGIIYGHVGMVDYIV 204


RESULT 7
Q97EB4
ID   Q97EB4      PRELIMINARY;      PRT;   273 AA.
AC   Q97EB4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Predicted transcriptional regulator, homolog of Bvg accessory
DE   factor.
GN   CAC3200.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
```

```
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007815; AAK81136.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30331 MW;  DE7B4D5923E72842 CRC64;

  Query Match             31.2%;  Score 335.5;  DB 16;  Length 273;
  Best Local Similarity   38.2%;  Pred. No. 1.5e-21;
  Matches   76;  Conservative   42;  Mismatches   74;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||: :||  :| :   : | ||  :: |  |:::: : :   | |  |   :
Db         14 LVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGVII 73

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :     :|     |:  || |   :: :|: | ||||||:| | ||  :|
Db         74 SSVVPNIMYSLEHMIRKYFKINPLVVGPG-IKTGINIKYDNPKEVGADRIVNAVAAHEIY 132

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                ||: :||||||||  | | | |||| ||||:  |: ||  : |||||:  |: : :
Db        133 --KRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRV--ELIKPA 188

Qy        184 PVIGRSTTTSLNHGFIFGF 202
              | ::| :|:  | ::|:
Db        189 YAICKNTISSIQSGIVYGY 207


RESULT 8
Q9KGH5
ID   Q9KGH5      PRELIMINARY;      PRT;   254 AA.
AC   Q9KGH5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein BH0086.
GN   BH0086.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001507; BAB03805.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   254 AA;  27907 MW;  0E3B557BA7DAC176 CRC64;

  Query Match             30.7%;  Score 330.5;  DB 16;  Length 254;
  Best Local Similarity   36.7%;  Pred. No. 3.8e-21;
  Matches   76;  Conservative   42;  Mismatches   82;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||  :|:  :  ::  : | |   :| |   :  :  ||||   |:   |
Db          3 LVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGIVI 62

Qy         66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   : : |::| :   ::  || |   |: :|: | ||||||:| | ||  ||
Db         63 SSVVPPMMFSLEQMCKKYFHVTPMIIGPG-IKTGLNIKYDNPKEVGADRIVNAVAAIELY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  : ||||||||:  : |   | ||:| ||:: |  ||  | :|||||  |: :
Db        122 GYP--AIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRI--EIAKPK 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:| :|  |:  |  :|: :  :||:
Db        178 QVVGTNTIDSMQSGIFYGYVSQVDGVV 204


RESULT 9
Q9A6Z1
ID   Q9A6Z1      PRELIMINARY;      PRT;   261 AA.
AC   Q9A6Z1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcriptional activator, putative, Baf family.
GN   CC1935.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005867; AAK23910.1; -.
DR   TIGR; CC1935; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   261 AA;  27965 MW;  C19E60D7B0714EF5 CRC64;

  Query Match             28.6%;  Score 307.5;  DB 16;  Length 261;
  Best Local Similarity   36.2%;  Pred. No. 4.1e-19;
  Matches   75;  Conservative   38;  Mismatches   91;  Indels    3;  Gaps    2;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |  : ||||    |   :: :  :  |:  :| |   : | ::|   |||   : | :
Db          4 LAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAVII 63

Qy         66 SSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||||     :|    ||   : |    :  : :| | |:|:| |||||| |  |  :||
Db         64 SSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYP 123

Qy        126 GPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
              ||  || :| |||||||  |   ||: ||:| ||:  |  ||    ||||||:::    :
Db        124 GP--LVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 181

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              ::|  | :::  |  :|: :: ||::|
Db        182 IVGTDTVSAMQSGVFWGYISLIEGLVA 208


RESULT 10
Q9X8N6
ID   Q9X8N6      PRELIMINARY;      PRT;   265 AA.
AC   Q9X8N6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   Hypothetical protein SCO3380.
GN   SCO3380 OR SCE94.31C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL049628; CAB40880.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   265 AA;  28203 MW;  E34FA2ABBA946969 CRC64;

  Query Match             26.0%;  Score 280;  DB 16;  Length 265;
  Best Local Similarity   35.5%;  Pred. No. 1.1e-16;
  Matches   76;  Conservative   36;  Mismatches   86;  Indels   16;  Gaps    8;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSI-----GLRLLEVLRHAGLGPADV 60
              |  |:|||:  :|:     ::  : : ||  :| | :     ||  :  |    ||
Db          3 LTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDGID 62

Qy         61 GACVASSVVPGVNPLIRRACERYL--YRKLLFAPG-DIAIPLDNRYERPAEVGADRLVAA 117
              |  : :: || |   :|    ||      :|  ||    :|:    : | ||||||:: |
Db         63 GIAICAT-VPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPI--LTDHPKEVGADRIINA 119

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRIS 176
               ||  || ||   : ||||||||||| |   | |:||:| ||:  |  ||  : |:| :|
Db        120 VAAVELYGGP--AIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKI- 176

Qy        177 LEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                ||     |||::|  ::  | ::|||   :||:
Db        177 -EVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVV 209


RESULT 11
Q9RX54
ID   Q9RX54      PRELIMINARY;      PRT;   262 AA.
AC   Q9RX54;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein DR0461.
GN   DR0461.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001905; AAF10040.1; -.
DR   TIGR; DR0461; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001220; Lectin_legB.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   262 AA;  27839 MW;  965EAAD2F78785A0 CRC64;

  Query Match             25.2%;  Score 271;  DB 16;  Length 262;
  Best Local Similarity   36.3%;  Pred. No. 6.5e-16;
  Matches   77;  Conservative   32;  Mismatches   87;  Indels   16;  Gaps    7;

Qy          6 LLFDIGNTNVKIGIAVETAVLT-SYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV-GAC 63
              |  |||||   :|:|  :  || :: : |:     | : |:|  :   ||   | :  |
Db          7 LAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAG---APIPRAA 63

Qy         64 VASSVVPGVNPLIRRACERYLYRKLLFAPG----DIAIPLDNRYERPAEVGADRLVAAYA 119
              | ||| | |       | :|:               |: : ||     |  ||||||   :
Db         64 VLSSVAPPVGENYALALKRHFMIDAFAVSAENLPDVTVELDT----PGSVGADRLCNLFG 119

Qy        120 ARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLE 178
              | :   |    | |||||:| || |  |  :|||::  |   || || :| ||||||:|:
Db        120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy        179 VEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                | :  ||::|  :|  | :||:| | :|:|
Db        180 APETA--IGKNTVHALQSGLVFGYAEMVDGLL 209


RESULT 12
Q9WZY5
ID   Q9WZY5      PRELIMINARY;      PRT;   246 AA.
AC   Q9WZY5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TM0883.
GN   TM0883.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
```

```
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001754; AAD35964.1; -.
DR   TIGR; TM0883; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   246 AA;  27154 MW;  9E0309AD462CF266 CRC64;

  Query Match             20.5%;  Score 221;  DB 16;  Length 246;
  Best Local Similarity   31.9%;  Pred. No. 1.5e-11;
  Matches   68;  Conservative   31;  Mismatches   92;  Indels   22;  Gaps    7;

Qy          5 FLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPA--DVGA 62
              :|| |:|||:    |  :      : | |   || | :   |     |. || |  ::
Db          2 YLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHL-----HPLLGDAMREIKG 56

Qy         63 CVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDN-----RYERPAEVGADRLVAA 117
              :|||| | :| |  ::|      |    | :   |       : |:||||||:
Db         57 IGVASVVPTQNTVIERFSQKY------FHISPIWVKAKNGCVKWNVKNPSEVGADRVANV 110

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISL 177
              .|  : | |   :: :| ||||| | |  |:| || | ||      :|   ||||| :
Db        111 VAFVKEY-GKNGII-IDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLV-- 166

Qy        178 EVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              ||:    |:|: |  ::  | : |     ||::
Db        167 EVKPADFVVGKDTEENIRLGVVNGSVYALEGII 199


RESULT 13
Q8RFE4
ID   Q8RFE4      PRELIMINARY;      PRT;   256 AA.
AC   Q8RFE4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bvg accessory factor.
GN   FN0761.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010586; AAL94957.1; -.
KW   Complete proteome.
SQ   SEQUENCE   256 AA;  28601 MW;  CDF681127F08914B CRC64;

  Query Match             20.4%;  Score 219.5;  DB 16;  Length 256;
  Best Local Similarity   29.9%;  Pred. No. 2.1e-11;
  Matches   64;  Conservative   35;  Mismatches   90;  Indels   25;  Gaps    6;

Qy          9 DIGNTNVKIGIAVETA-VLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVASS 67
              |||||::  ||      ::::: : |:   | |        : ::  :    | | : ||
Db          6 DIGNTHIVTGIYDNNGELISTFRIATNDKMTEDEYFSYFNNITKYNEISIKKVDAILISS 65

Qy         68 VVPGVNPLIRRACERY---------LYRKL--LFAPGDIAIPLDNRYERPAEVGADRLVA 116
              ||| :    :   :|         | :||   || |               ||||::
Db         66 VVPNIIITFQFFARKYFKVEATIVDLEKKLPFTFAKG----------INYTGFGADRIID 115

Qy        117 AYAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRIS 176
              | : ||  ::||  |||||||:| :: | |:|| | ||:  |  ||  ||||||:
Db        116 ITEAMQKYP-DKNLVIFDFGTATTYDVLKKGVYIGGGILPGIDMSINALYGNTAKLPRVK 174

Qy        177 LEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                  | |:|  |   :     ||:|   : ::
Db        175 FTT--PSSVLGTDTMKQIQAAIFFGYAGQIKHII 206


RESULT 14
Q9CD56
ID   Q9CD56      PRELIMINARY;      PRT;   274 AA.
AC   Q9CD56;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein ML0232.
GN   ML0232.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; AL583917; CAC29740.1; -.
DR   Leproma; ML0232; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   274 AA;  29421 MW;  1C2E735BDEC78765 CRC64;

  Query Match             19.5%;  Score 210.5;  DB 16;  Length 274;
  Best Local Similarity   31.2%;  Pred. No. 1.4e-10;
  Matches   67;  Conservative   41;  Mismatches   86;  Indels   21;  Gaps   10;

Qy          6 LLFDIGNTNVKIGI----AVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              |  |: ||:  :|:         |:  : : |:   | | : | ::: |   |     :
Db          3 LAIDVRNTHTVVGLLSGSKEHAKVVQQWRIRTESEVTADELAL-IIDGL--IGDDSERLA 59

Qy         62 ACVASSVVPGVNPLIRRACERY--LYRKLLFAPG---DIAIPLDNRYERPAEVGADRLVA 116
              | | || |   :|   ::|      :|  ||     | : :||    | ||||||:|
Db         60 GAAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDN----PKEVGADRIVN 115

Qy        117 AYAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRI 175
              ||   : |  ::| ||||::   | |   | :||| | |||  |: | ::|:| | |:
Db        116 CLAAFHKF-GQAAIV-VDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRV 173

Qy        176 SLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:    |:|::|   :  | :|||| : :|::
Db        174 --ELARPRSVVGKNTVECMQAGVVFGFAGLVDGLV 206


RESULT 15
O06282
ID   O06282      PRELIMINARY;      PRT;   272 AA.
AC   O06282;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 29.3 kDa protein (Transcriptional activator, putative,
DE   Baf family).
GN   RV3600C OR MTCY07H7B.22 OR MT3706.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z95557; CAB08944.1; -.
DR   EMBL; AE007170; AAK48063.1; -.
DR   TIGR; MT3706; -.
DR   TubercuList; Rv3600c; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   272 AA;  29304 MW;  5D70E6EF0F09AC8B CRC64;

  Query Match             19.5%;  Score 209.5;  DB 16;  Length 272;
  Best Local Similarity   30.3%;  Pred. No. 1.7e-10;
  Matches   66;  Conservative   41;  Mismatches   84;  Indels   27;  Gaps    9;

Qy          6 LLFDIGNTNVKIGI----AVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              |  |: ||:  :|:         |:  : : |:   | | : | :  ::   |     :
Db          3 LAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLI---GEDSERLT 59

Qy         62 ACVASSVVPGVNPLIRRACERY--LYRKLLFAPG---DIAIPLDNRYERPAEVGADRLV- 115
                | | || |   :|   ::|      :|  ||    | : :||    | ||||||:|
Db         60 GTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDN----PKEVGADRIVN 115

Qy        116 --AAYAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKL 172
                 |||   |    :: : ||||::   | |   | :||| | |||  |: | ::|:| |
Db        116 CLAAYDRFR-----KAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAAL 170

Qy        173 PRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:  |:     |:|::|   :  | :|||| : :|::
Db        171 RRV--ELARPRSVVGKNTVECMQAGAVFGFAGLVDGLV 206


Search completed: June 24, 2003, 22:16:20
Job time : 20.6485 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 4.26458 Seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453A-59
Perfect score: 1077
Sequence: 1 MTQHFLLFDIGNTNVKIGIA.....SLNHGFIFGFAAMTEGVLAA 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	338.5	31.4	233	1	YACB_BACSU	P37564 bacillus su
2	138	12.8	267	1	BAF_BORPE	Q45338 bordetella
3	86.5	8.0	458	1	TRKA_VIBAL	P39448 vibrio algi
4	84	7.8	227	1	GL13_ARATH	Q9m8x3 arabidopsis
5	83	7.7	407	1	HUTI_AGRRH	Q9kwe5 agrobacteri
6	82.5	7.7	485	1	HXKB_YEAST	P04807 saccharomyc
7	81.5	7.6	287	1	PSA3_YEAST	P21242 saccharomyc
8	81.5	7.6	1609	1	CTPI_MYCLE	O53114 mycobacteri
9	80	7.4	471	1	PPOX_MYXXA	P56601 myxococcus
10	78.5	7.3	455	1	NTPA_PEA	P52914 pisum sativ
11	78	7.2	271	1	TRPA_MYCIT	O68906 mycobacteri
12	78	7.2	351	1	LPXD_RHILO	Q98mc4 rhizobium l
13	78	7.2	1711	1	PTPO_RAT	Q64612 rattus norv
14	77.5	7.2	485	1	HXKA_YEAST	P04806 saccharomyc
15	77	7.1	2210	1	RRPL_EBOSM	Q66802 ebola virus
16	76.5	7.1	186	1	ATPD_RHOBL	P05437 rhodopseudo
17	76.5	7.1	251	1	ABCX_ANTSP	Q02856 antithamnio
18	76.5	7.1	266	1	TROB_TREPA	P96117 treponema p
19	76	7.1	444	1	PURA_SYNY3	P73290 synechocyst
20	76	7.1	495	1	VPN_BPMU	P08557 bacteriopha
21	76	7.1	1217	1	TOP1_ZYMMO	Q9x3x7 zymomonas m
22	75	7.0	336	1	YTCB_PSESQ	P27103 pseudomonas
23	75	7.0	354	1	CAHC_HUMAN	O43570 homo sapien
24	75	7.0	377	1	PROB_MEIRU	Q9zg98 meiothermus
25	75	7.0	392	1	EFT3_STRCO	P40175 streptomyce
26	75	7.0	478	1	Y4RH_RHISN	P55641 rhizobium s
27	74.5	6.9	213	1	Y767_MYCTU	P71822 mycobacteri
28	74.5	6.9	1011	1	UBAL_HUMAN	P41226 homo sapien
29	74.5	6.9	1053	1	RPOC_BROTH	P77839 brochothrix
30	74	6.9	330	1	PDXA_XYLFA	Q9pf39 xylella fas
31	74	6.9	351	1	MRAY_METTH	O26830 methanobact
32	74	6.9	389	1	EFT3_STRRA	P29544 streptomyce
33	74	6.9	409	1	HEM1_BRAJA	P08262 bradyrhizob
34	73.5	6.8	255	1	THYL_RAT	P01150 rattus norv
35	73.5	6.8	933	1	PRGR_HUMAN	P06401 homo sapien
36	73.5	6.8	3010	1	POLG_HCVJA	P26662 h genome po
37	73	6.8	317	1	RLA0_CERCA	Q9u3u0 ceratitis c
38	73	6.8	487	1	YK75_MYCTU	Q10683 mycobacteri
39	73	6.8	499	1	GLCD_ECOLI	P52075 escherichia
40	73	6.8	783	1	NQO3_THETH	Q56223 thermus the
41	73	6.8	1705	1	PTPO_MOUSE	P70289 mus musculu
42	72.5	6.7	227	1	COX2_HUMAN	P00403 homo sapien
43	72.5	6.7	284	1	LPXD_METCA	Q9aip8 methylococc
44	72.5	6.7	292	1	NIFH_HERSE	P77873 herbaspiril
45	72.5	6.7	344	1	PUR5_HAEIN	P43848 haemophilus

ALIGNMENTS

RESULT 1
YACB_BACSU
ID YACB_BACSU STANDARD; PRT; 233 AA.
AC P37564;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yacB.
GN YACB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=96051385; PubMed=7584024;
RA Ogasawara N., Nakai S., Yoshikawa H.;
RT "Systematic sequencing of the 180 kilobase region of the Bacillus
RT subtilis chromosome containing the replication origin.";
RL DNA Res. 1:1-14(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256(1997).
CC -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match           31.4%;  Score 338.5;  DB 1;  Length 233;
  Best Local Similarity   37.2%;  Pred. No. 2.5e-24;
  Matches   77;  Conservative   40;  Mismatches   83;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|:  :  :   : : |   :| |  |: |  :  |:||   :   :
Db          3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy         66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   : | | :| :    :  || :   |: :|: | ||||||:| | ||  ||
Db         63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  |: |||||||||:  : |   |:|| | ||:  |  || || |||||  |:
Db        122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              :||::| :::  | :||:    ||::
Db        178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204


RESULT 2
BAF_BORPE
ID   BAF_BORPE      STANDARD;      PRT;   267 AA.
AC   Q45338; Q45373;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Bvg accessory factor.
GN   BAF.
OS   Bordetella pertussis.
     Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
     Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RX   MEDLINE=95325323; PubMed=7601846;
RA   Deshazer D., Wood G.E., Friedman R.L.;
RT   "Identification of a Bordetella pertussis regulatory factor required
RT   for transcription of the pertussis toxin operon in Escherichia
RT   coli.";
RL   J. Bacteriol. 177:3801-3807(1995).
RN   [2]
RP   SEQUENCE OF 1-38 FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 239-267 FROM N.A.
RC   STRAIN=BP536;
RX   MEDLINE=96419162; PubMed=8821935;
RA   Allen A.G., Maskell D.J.;
RT   "The identification, cloning and mutagenesis of a genetic locus
RT   required for lipopolysaccharide biosynthesis in Bordetella
RT   pertussis.";
RL   Mol. Microbiol. 19:37-52(1996).
CC   -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC       A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC       RNA POLYMERASE.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U12020; AAA75361.1; -.
DR   EMBL; AF016461; AAC68834.1; -.
DR   EMBL; X90711; CAA62242.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Transcription regulation; Activator.
SQ   SEQUENCE   267 AA;  27845 MW;  336A615F67B57901 CRC64;

  Query Match           12.8%;  Score 138;  DB 1;  Length 267;
  Best Local Similarity   27.4%;  Pred. No. 1.2e-05;
  Matches   62;  Conservative   34;  Mismatches   78;  Indels   52;  Gaps   12;

Qy          6 LLFDIGNTNVKIG--------------IAVETAVLTSY--VLPTDPGQTTDSIGLRLLE 48
              :| | ||: :|:|                :| :   | :    | | | :   ::|: :
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNV-- 60

Qy         49 VLRHAGLGPADVGACVASSVVPGVNPLIRRACE-RYLYRKLLFAPGDIAIPLDNRYERPA 107
                 |||     |  :|:::  |        |: |:|  :       :|: | | |  |
Db         61 ----AGLAR---GEAIAATLRAG-------GCDIRWLRAQ------PLAMGLRNGYRNPD 100

Qy        108 EVGADR--LVAAYAARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGA 164
              ::||||   :    ||:    |  ||:  |||||| | :     : |||| ||     ||
Db        101 QLGADRWACMVGVLARQPSVHPPLLVA-SFGTATTLDTIGPDNVFPGGLILPGPAMMRGA 159

Qy        165 LSSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:  || || ::  :  | |:        |        | ||   | :
Db        160 LAYGTAHLP-LADGLVADYPIDTHQAIAS-------GIAAAQAGAI 197


RESULT 3
TRKA_VIBAL
ID   TRKA_VIBAL     STANDARD;      PRT;   458 AA.
AC   P39448;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Trk system potassium uptake protein trkA (K(+)-uptake protein trkA).
GN   TRKA.
OS   Vibrio alginolyticus.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=663;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=138-2;
RX   MEDLINE=95035102; PubMed=7948029;
RA   Nakamura T., Matsuba Y., Yamamuro N., Booth I.R., Unemoto T.;
RT   "Cloning and sequencing of a K+ transport gene (trk A) from the
RT   marine bacterium Vibrio alginolyticus.";
RL   Biochim. Biophys. Acta 1219:701-705(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=138-2;
RX   MEDLINE=98386520; PubMed=9720051;
RA   Nakamura T., Yamamuro N., Stumpe S., Unemoto T., Bakker E.P.;
RT   "Cloning of the trkAH gene cluster and characterization of the Trk
RT   K+-uptake system of Vibrio alginolyticus.";
RL   Microbiology 144:2281-2289(1998).
CC   -!- FUNCTION: PART OF A POTASSIUM TRANSPORT SYSTEM (BY SIMILARITY).
```

```
CC   -!- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 TRKA/RCK DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D28477; BAA05838.1; -.
DR   EMBL; D86411; BAA31228.1; -.
DR   InterPro; IPR000205; NAD_binding.
DR   InterPro; IPR000309; TrkA_Kuptake.
DR   InterPro; IPR003148; TrkA_N.
DR   Pfam; PF02080; TrkA-C; 2.
DR   Pfam; PF02254; TrkA-N; 2.
DR   PRINTS; PR00335; KUPTAKETRKA.
KW   Transport; Potassium transport; NAD; Repeat.
FT   DOMAIN       1    125       TRKA 1.
FT   DOMAIN     234    350       TRKA 2.
FT   NP_BIND      2     31       NAD (POTENTIAL).
FT   NP_BIND    234    262       NAD (POTENTIAL).
SQ   SEQUENCE   458 AA;  50125 MW;  309B2AC490C03472 CRC64;

  Query Match             8.0%;  Score 86.5;  DB 1;  Length 458;
  Best Local Similarity   27.0%;  Pred. No. 1.3;
  Matches   48;  Conservative   20;  Mismatches   67;  Indels   43;  Gaps   10;

Qy         40 DSIGLRLL-------EVLRHAGLGPAD------------VGAC-VASSVVPGVNPLIRRA 79
              |   ||::       :||  ||   ||            : || || ::    | : |
Db         41 DKYDLRVVNGHASHPDVLHEAGAQDADMLVAVTNTDETNMAACQVAFTLFNTPNRVARIR 100

Qy         80 CERYLYRK-LLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRL-YPGPRSLVSVDFGT 137
                ||  |  ||  |  |||:|:       :  : || :|  | : |||   :||
Db        101 SPEYLAEKEALFKSG--AIPVDHL------IAPEELVTSYIERLIQYPGALQVVSF---A 149

Qy        138 ATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEV----EEDSPVIGRSTT 191
                |   || || :          |||:    :| |   |      :  |:  : ||
Db        150 EQKVSLVAVKAYYGGPLV------GNALSALREHMPHIDTRVAAIFRQGRPIRPQGTT 201


RESULT 4
GL13_ARATH
ID   GL13_ARATH     STANDARD;      PRT;   227 AA.
AC   Q9M8X3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Germin-like protein subfamily 1 member 3 precursor.
GN   AT3G04170 OR T6K12.21.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=21016720; PubMed=11130713;
RA   Salanoubat M., Lemcke K., Rieger M., Ansorge W., Unseld M.,
RA   Fartmann B., Valle G., Bloecker H., Perez-Alonso M., Obermaier B.,
RA   Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA   De Simone V., Choisne N., Artiguenave F., Robert C., Brottier P.,
RA   Wincker P., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA   Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA   Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA   Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,
RA   Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simionati B.,
RA   Conrad A., Hornischer K., Kauer G., Loehnert T.-H., Nordsiek G.,
RA   Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA   Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA   Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA   de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA   Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA   Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA   Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA   Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA   Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA   Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA   Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA   Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA   Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA   Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA   Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT   "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 408:820-822(2000).
CC   -!- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO
CC       OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.
CC   -!- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY
CC       HEXAMER) (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Apoplast (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC016829; AAF26796.1; -.
DR   HSSP; P45850; 1FI2.
DR   InterPro; IPR001929; Germin.
DR   Pfam; PF01072; Germin; 1.
DR   PRINTS; PR00325; GERMIN.
DR   PROSITE; PS00725; GERMIN; 1.
KW   Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
KW   Multigene family.
FT   SIGNAL       1     24       POTENTIAL.
FT   CHAIN       25    227       GERMIN-LIKE PROTEIN SUBFAMILY 1 MEMBER 3.
FT   METAL      109    109       MANGANESE (BY SIMILARITY).
FT   METAL      111    111       MANGANESE (BY SIMILARITY).
FT   METAL      116    116       MANGANESE (BY SIMILARITY).
FT   METAL      160    160       MANGANESE (BY SIMILARITY).
FT   DISULFID    34     50       BY SIMILARITY.
FT   CARBOHYD   136    136       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   227 AA;  24424 MW;  C42948B9F92F9304 CRC64;

  Query Match             7.8%;  Score 84;  DB 1;  Length 227;
  Best Local Similarity   22.8%;  Pred. No. 1;
  Matches   44;  Conservative   23;  Mismatches   66;  Indels   60;  Gaps    8;

Qy         11 GNTNVKIGIAVETAVLTSYVLPTD-PGQTTDSIGLRLLEVLRHAGLGPADVGACVASSVV 69
              |||| ::|         |:| | : ||  |              |:| | :       :
Db         70 GNTNKRLG---------SFVNPANIPGLNT------------LGVGIARIDFAPGGLIP 107

Qy         70 PGVNPLIRRACE------------------RY-LYRKLLFAPGDIAIPLDNRYERPAEV 109
              | ::|   || |                  | |: |:|: |||: :       :  | :
Db        108 PHIHP---RASEIILVIKGKLLVGFVSSNDYNYTLFSKILY-PGDVFVHPIGLVQFHANI 163

Qy        110 GADRLVAAYAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRT 169
              |    ||  |      ||  |:    ||:    |              | :|: | ||
Db        164 GKTNAVAIGAVGSQNPGYISVGDAVFGSKPPID------------PKILAKAFALDINI 210

Qy        170 AKLPRISLEVEED 182
                :  |     ::|
Db        211 VRYLRKVFSPQDD 223


RESULT 5
HUTI_AGRRH
```

```
ID   HUTI_AGRRH     STANDARD;      PRT;   407 AA.
AC   Q9KWE5;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Imidazolonepropionase (EC 3.5.2.7) (Imidazolone-5-propionate
DE   hydrolase).
GN   HUTI OR RIORF39.
OS   Agrobacterium rhizogenes.
OG   Plasmid pRi1724.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Rhizobium.
OX   NCBI_TaxID=359;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF03-01724;
RA   Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA   Yoshida K.;
RT   "The complete nucleotide sequence of a Ri (root inducing) plasmid
RT   indicates its chimerical structure between Ti and Sym plasmids.";
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: 4-imidazolone-5-propanoate + H(2)O = N-
CC       formimino-L-glutamate.
CC   -!- PATHWAY: Histidine degradation; third step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE HUTI FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB039932; BAA97750.1; ALT_INIT.
DR   EMBL; AP002086; BAB16158.1; ALT_INIT.
DR   InterPro; IPR002604; ATZ_TRZ.
DR   Pfam; PF01685; ATZ_TRZ; 1.
DR   TIGRFAMs; TIGR01224; hutI; 1.
KW   Hydrolase; Histidine metabolism; Plasmid.
SQ   SEQUENCE   407 AA;  43511 MW;  A618F937E658D4B3 CRC64;

  Query Match             7.7%;  Score 83;  DB 1;  Length 407;
  Best Local Similarity   25.0%;  Pred. No. 2.5;
  Matches   36;  Conservative   15;  Mismatches   39;  Indels   54;  Gaps    6;

Qy        107 AEVGADRLVAAYAARRLYPGPRSLVSVDFGTATTFDC---------------------- 143
              ||:  | :: |   |  : || | :|:|   | | ||
Db         23 AEIINDAVIVAKEGRLTFVGPASALSIDDRDAETIDCGGRLITPGLVDCHTHIVFGGDRA 82

Qy        144 ------VEGGAY-----LGGLICPGVLSSAGALSSRT------AKLPRISLEVEEDSPVI 186
                   :||  |      ||    |::||  | :: |      | |||:   : |
Db         83 MEFEMRLEGSTYEEVARAGG----GIVSSVKATNALTVDQLVQASLPRVDTLLSE----- 133

Qy        187 GRSTTTSLNHGFIFGFAAMTEGVL 210
              | ||    :     |:    || |
Db        134 GVSTLEIKS-----GYGLTIEGEL 152


RESULT 6
HXKB_YEAST
ID   HXKB_YEAST     STANDARD;      PRT;   485 AA.
AC   P04807;
DT   13-AUG-1987 (Rel. 05, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Hexokinase B (EC 2.7.1.1) (Hexokinase PII).
GN   HXK2 OR HKB OR HEX1 OR YGL253W OR NRB486.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86120382; PubMed=3003701;
RA   Stachelek C., Stachelek J., Swan J., Botstein D., Konigsberg W.;
RT   "Identification, cloning and sequence determination of the genes
RT   specifying hexokinase A and B from yeast.";
RL   Nucleic Acids Res. 14:945-963(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86056943; PubMed=3905511;
RA   Froehlich K.-U., Entian K.-D., Mecke D.;
RT   "The primary structure of the yeast hexokinase PII gene (HXK2) which
RT   is responsible for glucose repression.";
RL   Gene 36:105-111(1985).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=97127827; PubMed=8972578;
RA   Coissac E., Maillier E., Robineau S., Netter P.;
RT   "Sequence of a 39,411 bp DNA fragment covering the left end of
RT   chromosome VII of Saccharomyces cerevisiae.";
RL   Yeast 12:1555-1562(1996).
RN   [4]
RP   SEQUENCE OF 1-246 FROM N.A.
RC   STRAIN=W303;
RX   MEDLINE=93311123; PubMed=8322518;
RA   Breitwieser W., Price C., Schuster T.;
RT   "Identification of a gene encoding a novel zinc finger protein in
RT   Saccharomyces cerevisiae.";
RL   Yeast 9:551-556(1993).
RN   [5]
RP   SEQUENCE OF 118-126; 175-184 AND 303-313.
RC   STRAIN=ATCC 38531 / Y41;
RX   MEDLINE=95255188; PubMed=7737086;
RA   Norbeck J., Blomberg A.;
RT   "Gene linkage of two-dimensional polyacrylamide gel electrophoresis
RT   resolved proteins from isogene families in Saccharomyces cerevisiae
RT   by microsequencing of in-gel trypsin generated peptides.";
RL   Electrophoresis 16:149-156(1995).
RN   [6]
RP   PHOSPHORYLATION OF SER-14.
RX   MEDLINE=94114477; PubMed=8286332;
RA   Kriegel T.M., Rush J., Vojtek A.B., Clifton D., Fraenkel D.G.;
RT   "In vivo phosphorylation site of hexokinase 2 in Saccharomyces
RT   cerevisiae.";
RL   Biochemistry 33:148-152(1994).
RN   [7]
RP   PHOSPHORYLATION OF SER-157.
RX   MEDLINE=97199316; PubMed=9047292;
RA   Heidrich K., Otto A., Behlke J., Rush J., Wenzel K.W., Kriegel T.;
RT   "Autophosphorylation-inactivation site of hexokinase 2 in
RT   Saccharomyces cerevisiae.";
RL   Biochemistry 36:1960-1964(1997).
RN   [8]
RP   SEQUENCE OF 1-18, AND PHOSPHORYLATION OF SER-14.
RX   MEDLINE=98384167; PubMed=9718324;
RA   Behlke J., Heidrich K., Naumann M., Mueller E.-C., Otto A., Reuter R.,
RA   Kriegel T.;
RT   "Hexokinase 2 from Saccharomyces cerevisiae: regulation of oligomeric
RT   structure by in vivo phosphorylation at serine-14.";
RL   Biochemistry 37:11989-11995(1998).
RN   [9]
RP   X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS).
RX   MEDLINE=78244654; PubMed=355643;
RA   Anderson C.M., Stenkamp R.E., Steitz T.A.;
RT   "Sequencing a protein by X-ray crystallography. II. Refinement of
RT   yeast hexokinase B co-ordinates and sequence at 2.1-A resolution.";
RL   J. Mol. Biol. 123:15-33(1978).
CC   -!- FUNCTION: MAIN GLUCOSE PHOSPHORYLATING ENZYME. MAY PLAY A
CC       REGULATORY ROLE IN BOTH INDUCTION AND REPRESSION OF GENE
CC       EXPRESSION BY GLUCOSE.
CC   -!- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.
CC   -!- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
```

```
CC       INHIBITION BY ATP.
CC   -!- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
CC       ISOENZYMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
CC   -!- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/H/HK.html".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X03483; CAA27203.1; -.
DR   EMBL; M11181; AAA34697.1; -.
DR   EMBL; X94357; CAA64134.1; -.
DR   EMBL; Z72775; CAA96973.1; -.
DR   EMBL; X67787; CAA48003.1; -.
DR   PIR; B23523; KIBYHB.
DR   PIR; S28555; S28555.
DR   PDB; 2YHX; 15-JUL-92.
DR   SWISS-2DPAGE; P04807; YEAST.
DR   SGD; S0003222; HXK2.
DR   InterPro; IPR001312; Hexokinase.
DR   Pfam; PF00349; hexokinase; 1.
DR   Pfam; PF03727; hexokinase2; 1.
DR   PRINTS; PR00475; HEXOKINASE.
DR   ProDom; PD001109; Hexokinase; 1.
DR   PROSITE; PS00378; HEXOKINASES; 1.
KW   Transferase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;
KW   3D-structure; Phosphorylation.
FT   INIT_MET      0      0
FT   BINDING     110    110       ATP (BY SIMILARITY).
FT   DOMAIN      151    177       GLUCOSE-BINDING (POTENTIAL).
FT   MOD_RES      14     14       PHOSPHORYLATION.
FT   MOD_RES     157    157       PHOSPHORYLATION.
FT   CONFLICT     28     28       N -> I (IN REF. 1).
FT   CONFLICT     32     32       I -> N (IN REF. 2).
FT   CONFLICT     60     60       G -> V (IN REF. 1).
FT   CONFLICT    196    196       T -> S (IN REF. 1).
FT   CONFLICT    420    421       YN -> ST (IN REF. 2).
FT   CONFLICT    443    444       TS -> PH (IN REF. 2).
FT   CONFLICT    452    452       I -> V (IN REF. 2).
FT   CONFLICT    461    461       A -> P (IN REF. 2).
FT   HELIX        23     33
FT   HELIX        37     55
FT   STRAND       65     66
FT   STRAND       79     86
FT   STRAND       90     99
FT   TURN        100    101
FT   STRAND      102    110
FT   STRAND      112    112
FT   TURN        115    118
FT   TURN        123    123
FT   HELIX       124    141
FT   TURN        142    142
FT   STRAND      150    155
FT   STRAND      164    164
FT   TURN        165    166
FT   STRAND      169    169
FT   TURN        175    176
FT   STRAND      187    187
FT   HELIX       188    199
FT   STRAND      202    208
FT   HELIX       210    221
FT   TURN        223    224
FT   STRAND      225    231
FT   STRAND      235    241
FT   HELIX       243    245
FT   TURN        251    252
FT   STRAND      263    266
FT   TURN        270    275
FT   HELIX       283    291
FT   HELIX       299    304
FT   HELIX       306    322
FT   TURN        323    324
FT   TURN        338    339
FT   TURN        343    343
FT   HELIX       344    351
FT   HELIX       358    368
FT   TURN        369    369
FT   HELIX       374    395
FT   TURN        396    396
FT   HELIX       397    406
FT   TURN        407    407
FT   STRAND      411    416
FT   TURN        418    421
FT   TURN        424    425
FT   HELIX       426    438
FT   HELIX       445    447
FT   STRAND      450    454
FT   TURN        458    460
FT   HELIX       461    475
SQ   SEQUENCE   485 AA;  53811 MW;  E4A7EC796125324E CRC64;

  Query Match             7.7%;  Score 82.5;  DB 1;  Length 485;
  Best Local Similarity  23.2%;  Pred. No. 3.3;
  Matches   57;  Conservative   30;  Mismatches   94;  Indels   65;  Gaps   13;

Qy          5 FLLFDIGNTNVKI------GIAVETAVLTSYVLPTDPGQTT----------DSIGLRLL 47
              ||  |:| ||:::      |        : | || |  :||           ||:   :
Db         81 FLAIDLGGTNLRVVLVKLGGDRTFDTTQSKYRLP-DAMRTTQNPDELWEFIADSLKAFID 139

Qy         48 EVLRHAGLGPADVGACVASSVVPGVNPLIRRACERYLYRKLLFAPG----DIAIPLDNRY 103
              |       | :|     |     | :     :|  : |   |     |:   |  :
Db        140 EQFPQGISEPIPLG--FTFSFPASQNKINEGILQR--WTKGFDIPNIENHDVVPMLQKQI 195

Qy        104 ER---PAEVGA------DRLVAAYAARRLYPGPRSLVSVDFGTATTFDCVEGGAY----- 149
               :   | || |      |||:|     |  | : : | |||          |||
Db        196 TKRNIPIEVVALINDTTGTLVASY-----YTDPETKMGVIFGTGV------NGAYYDVCS 244

Qy        150 ----LGGLICPGVLSSA--------GALSSRTAKLPR--ISLEVEEDSPVIGRSTTTSLN 195
                  | | :   :  ||        |:  :    |||    : ::|:|| |: |   ::
Db        245 DIEKLQGKLSDDIPPSAPMAINCEYGSFDNEHVVLPRTKYDITIDEESPRPGQQTFEKMS 304

Qy        196 HGFIFG 201
                |:  |
Db        305 SGYYLG 310


RESULT 7
PSA3_YEAST
ID   PSA3_YEAST     STANDARD;      PRT;   287 AA.
AC   P21242;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Proteasome component C1 (EC 3.4.25.1) (Macropain subunit C1)
DE   (Proteinase YSCE subunit 1) (Multicatalytic endopeptidase complex
DE   subunit C1).
GN   PRS1 OR PRC1 OR PRE10 OR YOR362C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 105-134; 179-194 AND 199-208.
RX   MEDLINE=90375535; PubMed=1697860;
RA   Fujiwara T., Tanaka K., Orino E., Yoshimura T., Kumatori A.,
RA   Tamura T., Chung C.H., Nakai T., Yamaguchi K., Shin S.,
RA   Kakizuka A., Nakanishi S., Ichihara A.;
RT   "Proteasomes are essential for yeast proliferation. cDNA cloning and
```

```
RT   gene disruption of two major subunits.";
RL   J. Biol. Chem. 265:16604-16613(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Delius H., Hebling U., Hofmann B.;
RL   Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 4-247.
RX   MEDLINE=97242404; PubMed=9087403;
RA   Groll M., Ditzel L., Lowe J., Stock D., Bochtler M., Bartunik H.D.,
RA   Huber R.;
RT   "Structure of 20S proteasome from yeast at 2.4-A resolution.";
RL   Nature 386:463-471(1997).
CC   -!- FUNCTION: THE PROTEASOME IS A MULTICATALYTIC PROTEINASE COMPLEX
CC       WHICH IS CHARACTERIZED BY ITS ABILITY TO CLEAVE PEPTIDES WITH ARG,
CC       PHE, TYR, LEU, AND GLU ADJACENT TO THE LEAVING GROUP AT NEUTRAL OR
CC       SLIGHTLY BASIC PH. THE PROTEASOME HAS AN ATP-DEPENDENT PROTEOLYTIC
CC       ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: Cleavage at peptide bonds with very broad
CC       specificity.
CC   -!- PATHWAY: Involved in an ATP/ubiquitin-dependent non-lysosomal
CC       proteolytic pathway.
CC   -!- SUBUNIT: The yeast proteasome seems to be composed of 14 different
CC       subunits which form a highly ordered ring-shaped structure.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC   -!- PTM: ALPHA AND BETA FORMS EXIST. THEY ARE PROBABLY PRODUCTS OF THE
CC       SAME GENE WITH DIFFERENT POSTTRANSLATIONAL MODIFICATIONS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY T1A.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M55436; AAA35227.1; -.
DR   EMBL; Z75270; CAA99691.1; -.
DR   PIR; S11182; SNBYC1.
DR   PDB; 1RYP; 15-APR-98.
DR   MEROPS; T01.977; -.
DR   SGD; S0005889; PRE10.
DR   InterPro; IPR000426; Proteasome_A.
DR   InterPro; IPR001353; Protsme_protease.
DR   Pfam; PF00227; proteasome; 1.
DR   PROSITE; PS00388; PROTEASOME_A; 1.
KW   Proteasome; Hydrolase; Protease; 3D-structure.
FT   INIT_MET      0      0
FT   MOD_RES       1      1       BLOCKED (POTENTIAL).
SQ   SEQUENCE   287 AA;  31405 MW;  0AEE8B0E99063E54 CRC64;

  Query Match             7.6%;  Score 81.5;  DB 1;  Length 287;
  Best Local Similarity   27.2%;  Pred. No. 2.3;
  Matches   41;  Conservative   16;  Mismatches   51;  Indels   43;  Gaps    8;

Qy         10 IGNTNVKIGI--------AVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              : |    |||        |||  : :  ::|           :::  | || |
Db         30 VENGTTSIGIKCNDGVVFAVEKLITSKLLVPQK--------NVKIQVVDRHIG------- 74

Qy         62 ACVASSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120
                || | ::|    |: |  |    ::||   |  | ||             ||||    |
Db         75 -CVYSGLIPDGRHLVNRGREEAASFKKLYKTP--IPIP----------AFADRLGQYVQA 121

Qy        121 RRLYPGPRSLVSVDFGTATTFDCVE-GGAYL 150
                ||   |      || :| |  |:  ||:|
Db        122 HTLYNSVR-----PFGVSTIFGGVDKNGAHL 147


RESULT 8
CTPI_MYCLE
ID   CTPI_MYCLE     STANDARD;      PRT;  1609 AA.
AC   O53114;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable cation-transporting ATPase I (EC 3.6.3.-).
GN   CTPI OR ML2671 OR MLCB1913.02.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC       (E1-E2 ATPASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL022118; CAA17934.1; -.
DR   EMBL; AL583926; CAC32203.1; -.
DR   Leproma; ML2671; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR001454; Hlgnase/hydrlase.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW   Complete proteome.
FT   TRANSMEM     30     50       POTENTIAL.
FT   TRANSMEM    176    196       POTENTIAL.
FT   TRANSMEM    238    258       POTENTIAL.
FT   TRANSMEM    357    377       POTENTIAL.
FT   TRANSMEM    641    661       POTENTIAL.
FT   TRANSMEM    673    693       POTENTIAL.
FT   TRANSMEM    778    798       POTENTIAL.
FT   TRANSMEM    921    941       POTENTIAL.
FT   TRANSMEM    969    989       POTENTIAL.
FT   TRANSMEM    997   1017       POTENTIAL.
FT   TRANSMEM   1396   1416       POTENTIAL.
FT   TRANSMEM   1426   1446       POTENTIAL.
FT   TRANSMEM   1542   1562       POTENTIAL.
FT   TRANSMEM   1573   1593       POTENTIAL.
FT   MOD_RES    1053   1053       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL      1335   1335       MAGNESIUM (BY SIMILARITY).
FT   METAL      1339   1339       MAGNESIUM (BY SIMILARITY).
SQ   SEQUENCE   1609 AA;  166871 MW;  58FA2079905E3995 CRC64;

  Query Match             7.6%;  Score 81.5;  DB 1;  Length 1609;
  Best Local Similarity   24.7%;  Pred. No. 16;
  Matches   56;  Conservative   29;  Mismatches   69;  Indels   73;  Gaps   14;

Qy         15 VKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVASSVVPGVNP 74
              | :|  |:      | :|    ||    | |:  ||||   |  |: | | ::  ||  |
```

```
Db        510 VDVGWEVD-----PYAIPL--LQTAHRTGARV--VLRHVA-GTEDLSASVGATHPPG-TP 558

Qy         75 LIRRACERYLYRKL----------------------LFAPGDIAIPLDNRYER---PA 107
              |::     | |:|                       |  |: : ||: :     |
Db        559 LLK------LVRELRTDRGPVLLITAVHRDFASTDTLAALAIADVGVALDDPHAATPWTA 612

Qy        108 EV--GADRLVAAYAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGAL 165
              ::  | | | ||       |  ||       : ::    :||  | ||:   : :|||:
Db        613 DIITGTD-LAAAVRILSALPVARS------ASESSVHLAQGGTTLAGLLL--ITASAGSK 663

Qy        166 SSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212
              |:    | |       |||   : |           |:  ||::|
Db        664 SASPITLRRWF------SPVNAAAAT-----------ALVTGVVSA 692


RESULT 9
PPOX_MYXXA
ID   PPOX_MYXXA     STANDARD;      PRT;   471 AA.
AC   P56601;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Protoporphyrinogen oxidase (EC 1.3.3.4) (PPO).
GN   HEMY.
OS   Myxococcus xanthus.
OC   Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC   Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX   NCBI_TaxID=34;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92192453; PubMed=1372277;
RA   Li S., Lee B.U., Shimkets L.J.;
RT   "csgA expression entrains Myxococcus xanthus development.";
RL   Genes Dev. 6:401-410(1992).
RN   [2]
RP   REVISION, AND CHARACTERIZATION.
RX   MEDLINE=96224079; PubMed=8621504;
RA   Dailey H.A., Dailey T.A.;
RT   "Protoporphyrinogen oxidase of Myxococcus xanthus. Expression,
RT   purification, and characterization of the cloned enzyme.";
RL   J. Biol. Chem. 271:8714-8718(1996).
CC   -!- FUNCTION: CATALYZES THE 6-ELECTRON OXIDATION OF PROTOPORPHYRINOGEN
CC       IX TO FORM PROTOPORPHYRIN IX. MAY HAVE AN ESSENTIAL EXTRACELLULAR
CC       FUNCTION DURING DEVELOPMENT POSSIBLY AS A PHEROMONE.
CC   -!- CATALYTIC ACTIVITY: Protoporphyrinogen-IX + O(2) = protoporphyrin-
CC       IX + H(2)O(2).
CC   -!- COFACTOR: CONTAINS ONE FAD PER HOMODIMER.
CC   -!- PATHWAY: PENULTIMATE STEP IN HEME AND PORPHYRIN BIOSYNTHESIS.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO THE PROTOPORPHYRINOGEN OXIDASE FAMILY.
CC   -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 127.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M73709; -; NOT_ANNOTATED_CDS.
DR   EMBL; AF098938; AAD13609.1; -.
DR   InterPro; IPR002937; Amino_oxidase.
DR   InterPro; IPR000205; NAD_binding.
DR   InterPro; IPR004572; Proto_IX_ox.
DR   Pfam; PF01593; Amino_oxidase; 1.
DR   TIGRFAMs; TIGR00562; proto_IX_ox; 1.
KW   Porphyrin biosynthesis; Oxidoreductase; Flavoprotein; FAD;
KW   Heme biosynthesis.
FT   NP_BIND      16     21       FAD (POTENTIAL).
SQ   SEQUENCE   471 AA;  49387 MW;  2E9909D464F81515 CRC64;

  Query Match             7.4%;  Score 80;  DB 1;  Length 471;
  Best Local Similarity   22.6%;  Pred. No. 5.5;
  Matches   44;  Conservative   26;  Mismatches   69;  Indels   56;  Gaps    9;

Qy         13 TNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVASSV---V 69
              :: ::| || |  |  |::  |   |                ||      | ::  :
Db         40 SSARLGGAVGTHALAGYLVEQGPNSFLDR-------------EPATRALAAALNLEGRI 85

Qy         70 PGVNPLIRRACERYLYRK----------LLFAPGDIAIPLDNR-------YERPAEVGAD 112
                :|  :|   ||:| :            |   || :||  |       : | |  | |
Db         86 RAADPAAKR---RYVYTRGRLRSVPASPPAFLASDI-LPLGARLRVAGELFSRRAPEGVD 141

Qy        113 RLVAAYAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKL 172
              :||:  | |  | |:       |    | |: | | | :    | ::   |
Db        142 ESLAAFGRRHL--GHRA-------TQVLLDAVQTGIYAGDVEQLSVAATFPML------- 185

Qy        173 PRISLEVEEDSPVIG 187
              : :| |  | ::|
Db        186 --VKMEREHRSLILG 198


RESULT 10
NTPA_PEA
ID   NTPA_PEA       STANDARD;      PRT;   455 AA.
AC   P52914;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Nucleoside-triphosphatase (EC 3.6.1.15) (Nucleoside triphosphate
DE   phosphohydrolase) (NTPase) (Apyrase).
OS   Pisum sativum (Garden pea).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Pisum.
OX   NCBI_TaxID=3888;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Alaska; TISSUE=Plumule;
RX   MEDLINE=96197404; PubMed=8616230;
RA   Hsieh H., Tong C.G., Thomas C., Roux S.J.;
RT   "Light-modulated abundance of an mRNA encoding a calmodulin-regulated,
RT   chromatin-associated NTPase in pea.";
RL   Plant Mol. Biol. 30:135-147(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Alaska; TISSUE=Stem;
RA   Shibata K., Abe S., Davies E.;
RT   "Structure of the coding region and mRNA variants of the apyrase from
RT   Pisum sativum.";
RL   Acta Physiol. Plant. 20:3-13(2001).
CC   -!- FUNCTION: MIGHT BE INVOLVED IN RNA TRANSPORT OUT OF NUCLEI.
CC   -!- CATALYTIC ACTIVITY: NTP + H(2)O = NDP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE GDA1 / CD39 NTPASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z32743; CAA83655.1; -.
DR   EMBL; AB027614; BAB18900.1; -.
DR   InterPro; IPR000407; GDA1_CD39_NTPase.
DR   Pfam; PF01150; GDA1_CD39; 1.
DR   PROSITE; PS01238; GDA1_CD39_NTPASE; 1.
KW   Hydrolase; Nuclear protein.
SQ   SEQUENCE   455 AA;  50072 MW;  50FDF0023ABC4299 CRC64;

  Query Match             7.3%;  Score 78.5;  DB 1;  Length 455;
```

```
  Best Local Similarity   22.0%;  Pred. No. 7.3;
  Matches   54;  Conservative   37;  Mismatches   62;  Indels   93;  Gaps   14;

Qy          5 FLLFD---------IGN---TNVKIGIAVETAVLTSYVLPTDPGQTTDSIGL----RLLE 48
              ||||          :||   |: || :  |   ::|| :  | | |   | :    : |:
Db         10 FLLFSMPAITSSQYLGNNLLTSRKIFLKQEE--ISSYAVVFDAGSTGSRIHVYHFNQNLD 67

Qy         49 VLRHAGLGPADVGACVASSVVPGVN--------------PLIRRACERYLYRKLLFAPGD 94
              :| | | |         : : ||::              ||: :|              :
Db         68 LL-HIGKGVE-----YYNKITPGLSSYANNPEQAAKSLIPLLEQA-------------E 107

Qy         95 IAIPLDNRYERPAEVGAD---RLVAAYAARRLYPGPRSLVS------VDFGTATTFDCVE 145
                 :| | : : |  :||    ||:   |: ::    | ::|      |     :  |  :
Db        108 DVVPDDLQPKTPVRLGATAGLRLLNGDASEKILQSVRDMLSNRSTFNVQPDAVSIIDGTQ 167

Qy        146 GGAY-----------------------LGGLICPGVLSSAGALSSRTAK-LPRISLEVE 180
               |:|                      |||    | :  | |:| :||| |:::   :
Db        168 EGSYLWVTVNYALGNLGKKYTKTVGVIDLGG----GSVQMAYAVSKKTAKNAPKVA---D 220

Qy        181 EDSPVI 186
                | | |
          221 GDDPYI 226
```

```
RESULT 11
TRPA_MYCIT
ID   TRPA_MYCIT     STANDARD;      PRT;   271 AA.
AC   O68906;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tryptophan synthase alpha chain (EC 4.2.1.20).
GN   TRPA.
OS   Mycobacterium intracellulare.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1767;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Batty;
RA   Alavi M.R., Rouse D.A., Morris S.L.;
RT   "Nucleotide sequence and functional analysis of the tryptophan
RT   synthase genes of Mycobacterium intracellulare.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE ALPHA SUBUNIT IS RESPONSIBLE FOR THE ALDOL CLEAVAGE
CC       OF INDOLEGLYCEROL PHOSPHATE TO INDOLE AND GLYCERALDEHYDE 3-
CC       PHOSPHATE.
CC   -!- CATALYTIC ACTIVITY: L-serine + 1-(indol-3-yl)glycerol 3-phosphate
CC       = L-tryptophan + glyceraldehyde 3-phosphate + H(2)O.
CC   -!- PATHWAY: Tryptophan biosynthesis; fifth (last) step.
CC   -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRPA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF057042; AAC17135.1; -.
DR   HSSP; P00929; 2TYS.
DR   InterPro; IPR003009; FMN_enzyme.
DR   InterPro; IPR002028; Trp_synthaseA.
DR   Pfam; PF00290; trp_syntA; 1.
DR   ProDom; PD001535; Trp_synthaseA; 1.
DR   TIGRFAMs; TIGR00262; trpA; 1.
DR   PROSITE; PS00167; TRP_SYNTHASE_ALPHA; 1.
KW   Tryptophan biosynthesis; Lyase.
SQ   SEQUENCE   271 AA;  27804 MW;  BA0E0C5BC0788318 CRC64;
```

```
  Query Match             7.2%;  Score 78;  DB 1;  Length 271;
  Best Local Similarity   26.9%;  Pred. No. 4.5;
  Matches   52;  Conservative   23;  Mismatches   88;  Indels   30;  Gaps   12;

Qy         24 AVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVASSVVPGVN-PLIRRACER 82
              | |  | |||      | |:    :  |  :|    :||    :     ||:: | | :| |
Db         24 AALIGY-LPTGYPDVTTSV--HAMVALVESGCDIVEVGVPYSD---PGMDGPTIVKATEA 77

Qy         83 YLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYPGPRSLVSVDFGTATTFD 142
                 |:       | : :      :  | | |      :|    |  |     :|   |  |
Db         78 ALH-------GGVRV-----RDTLAAVEAISAAGGHAVVMTYWNPVLRYGID---AFARD 122

Qy        143 CVEGGAYLGGLICPGVL-SSAG--ALSSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFI 199
                 | |  ||| | ::   ||    :|   :| || |     :|    :  |   : ||:
Db        123 LASAGGY--GLITPDLIPDEAGQWLAASEEHRLDRIFLVAPSSTPQ-RLAMTVEASRGFV 179

Qy        200 FGFAAMTEGVLAA 212
                :|| | ||  |
Db        180 --YAASTMGVTGA 190
```

```
RESULT 12
LPXD_RHILO
ID   LPXD_RHILO     STANDARD;      PRT;   351 AA.
AC   Q98MC4;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase
DE   (EC 2.3.1.-).
GN   LPXD OR MLL0635.
OS   Rhizobium loti (Mesorhizobium loti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Phyllobacteriaceae; Mesorhizobium.
OX   NCBI_TaxID=381;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MAFF303099;
RX   MEDLINE=21082930; PubMed=11214968;
RA   Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA   Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA   Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA   Takeuchi C., Yamada M., Tabata S.;
RT   "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT   Mesorhizobium loti.";
RL   DNA Res. 7:331-338(2000).
CC   -!- CATALYTIC ACTIVITY: UDP-3-O-(3-hydroxytetradecanoyl)glucosamine +
CC       (R)-3-hydroxytetradecanoyl-[acyl-carrier protein] = UDP-2,3-bis(3-
CC       hydroxytetradecanoyl)glucosamine + [acyl-carrier protein].
CC   -!- PATHWAY: Lipid A biosynthesis; third step.
CC   -!- SIMILARITY: BELONGS TO THE TRANSFERASE HEXAPEPTIDE REPEAT FAMILY.
CC       LPXD SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AP002995; BAB48189.1; -.
DR   InterPro; IPR001451; Hexapep_transf.
DR   Pfam; PF00132; hexapep; 7.
DR   PROSITE; PS00101; HEXAPEP_TRANSFERASES; FALSE_NEG.
KW   Transferase; Acyltransferase; Lipid A biosynthesis; Lipid synthesis;
KW   Repeat; Complete proteome.
SQ   SEQUENCE   351 AA;  35584 MW;  7D865F67C7C0570F CRC64;
```

```
  Query Match             7.2%;  Score 78;  DB 1;  Length 351;
  Best Local Similarity   26.1%;  Pred. No. 6.1;
  Matches   58;  Conservative   25;  Mismatches   79;  Indels   60;  Gaps   14;
```

```
Qy         20 AVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVASSVVPG-----VNP 74
              | | | ||  ||  | ||:  ||            | ||  |   |  | |     : |
Db         15 AAEVANLTGSVL-VDSGQSDVSI----------EALAPASEGGANALVFVDGRRNSALMP 63

Qy         75 LIRRA---CERYLYRKLLFAPGDIAIPLDNRYERPAEVGA--DRLVAAYAARRLYPGPRS 129
               :| |   |       |   ||  :|: :   : || :  |   ||:   ||    |||
Db         64 SLRAAAVLCPAEFANK---APSSVAVLI---HPRPQQAFALVGRLLFPTAAT---PGP-- 112

Qy        130 LVSVDFGTATTFDC-----VEGGAYL--GGLICPGVLSSAGALSSRTAKLPRISLEVEED 182
               :: : | :          || || : | :| |||   :| : :  | : : | ::  |
Db        113 -MTSETGVSAHAHIDPTAHVEAGAIIEAGVVIGPGVSIGSGTVIAPNAVIGQ-SCQIGRD 170

Qy        183 SPVIGRSTTTSLNHGFI----------------FGFAAMTEG 208
               |       |: :  |                |||    :|
Db        171 GYV---GPGASIQYALIGNRVIIHGGARIGQDGFGFVGGAKG 209


RESULT 13
PTPO_RAT
ID   PTPO_RAT       STANDARD;      PRT;  1711 AA.
AC   Q64612;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Osteotesticular protein tyrosine phosphatase precursor (EC 3.1.3.48)
DE   (OST-PTP).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Osteosarcoma;
RX   MEDLINE=95074080; PubMed=7527035;
RA   Mauro L.J., Olmsted E.A., Skrobacz B.M., Mourey R.J., Davis A.R.,
RA   Dixon J.E.;
RT   "Identification of a hormonally regulated protein tyrosine
RT   phosphatase associated with bone and testicular differentiation.";
RL   J. Biol. Chem. 269:30659-30667(1994).
CC   -!- FUNCTION: MAY FUNCTION IN SIGNALING PATHWAYS DURING BONE
CC       REMODELING, AS WELL AS SERVE A BROADER ROLE IN CELL INTERACTIONS
CC       ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS. OPTIMAL PH FOR
CC       PHOSPHATASE ACTIVITY IS 5.6.
CC       ASSOCIATED WITH DIFFERENTIATION IN BONE AND TESTIS.
CC   -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC       tyrosine + phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
CC       produced by alternative splicing. A presumed alternate transcript
CC       of 4.8-5.0 kilobases, which may lack PTP domains, is present in
CC       proliferating osteoblasts, but not detectable at other stages.
CC   -!- TISSUE SPECIFICITY: BONE AND TESTIS. IN THE LATTER, RESTRICTED TO
CC       THE BASAL PORTION OF THE SEMINIFEROUS TUBULE.
CC   -!- DEVELOPMENTAL STAGE: UP-REGULATED IN DIFFERENTIATING CULTURES OF
CC       PRIMARY OSTEOBLASTS AND DOWN-REGULATED IN LATE STAGE MINERALIZING
CC       CULTURES. IN TESTIS, EXPRESSION IS HIGHEST BETWEEN STAGES I AND
CC       VII WHEN MATURING SPERMATIDS REMAIN BURIED WITHIN THE SERTOLI
CC       EPITHELIUM.
CC   -!- INDUCTION: BY PARATHYROID HORMONE AND CYCLIC AMP ANALOGS.
CC   -!- PTM: THE CYTOPLASMIC DOMAIN CONTAINS POTENTIAL PHOSPHORYLATION
CC       SITES.
CC   -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 10 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L36884; AAA63911.1; -.
DR   HSSP; P18052; 1YFO.
DR   InterPro; IPR000340; DS_phosphatase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR000387; TYR_phosphatase.
DR   InterPro; IPR000242; Tyr_PP.
DR   Pfam; PF00041; fn3; 7.
DR   Pfam; PF00102; Y_phosphatase; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PRINTS; PR00700; PRTYPHPHTASE.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00194; PTPc; 1.
DR   SMART; SM00012; PTPc_DSPc; 1.
DR   PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR   PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR   PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 2.
KW   Hydrolase; Transmembrane; Repeat; Signal; Glycoprotein.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN        18   1711       OSTEOTESTICULAR PROTEIN TYROSINE
FT                                PHOSPHATASE.
FT   DOMAIN       18   1074       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1075   1095       POTENTIAL.
FT   DOMAIN     1096   1711       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       32    124       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      125    215       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      216    303       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      304    392       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      393    470       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      471    562       FIBRONECTIN TYPE-III 6.
FT   DOMAIN      563    652       FIBRONECTIN TYPE-III 7.
FT   DOMAIN      653    741       FIBRONECTIN TYPE-III 8.
FT   DOMAIN      742    830       FIBRONECTIN TYPE-III 9.
FT   DOMAIN      831    921       FIBRONECTIN TYPE-III 10.
FT   DOMAIN     1150   1418       PROTEIN-TYROSINE PHOSPHATASE 1.
FT   DOMAIN     1419   1711       PROTEIN-TYROSINE PHOSPHATASE 2.
FT   ACT_SITE   1350   1350       BY SIMILARITY.
FT   CARBOHYD     42     42       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     89     89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    117    117       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    174    174       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    239    239       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    259    259       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    431    431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    570    570       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    620    620       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    649    649       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    663    663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    737    737       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    851    851       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    882    882       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    970    970       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    982    982       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1711 AA;  187292 MW;  0F04D2D1A47A18A0 CRC64;

  Query Match             7.2%;  Score 78;  DB 1;  Length 1711;
  Best Local Similarity   23.2%;  Pred. No. 36;
  Matches   62;  Conservative   25;  Mismatches   80;  Indels  100;  Gaps   15;

Qy          4 HFLLFD-IGNTN----VKIGIAVETAV-----------------------LTSYVLPT 33
              | :| | :| ||    :: |::  | :                         | :  |:
Db        248 HLMLTDLLGGTNTTAVIRQGVSTHTFLHLSPGTPHELKICASAGPHQIWGPSATEWTYPS 307

Qy         34 DPGQTT-------------DSIGLRLLEVLRHAG-------LGPADVGACVASSVVPGVN 73
                |                :| |   ||: :|       ||| :  |     : || .
Db        308 YPSDLVLTPLRNELWASWKAGLGARDGYVLKLSGPMESTSTLGPEECNAVFPGPLPPG-- 365

Qy         74 PLIRRACERYLYRKLLFAPGDIAIP----LDNRYERPAEVGADRL------VAAYAARR- 122
                      |  |:|  | |  :     |       | ||   ||       :    ||
Db        366 -------HYTLQLKVLAGPYDAWVEGSTWLAESAALPREVPGARLWLDGLEASKQPGRRA 418

Qy        123 -LY----PGPRSLVSVDFGTATTFDCVEGGAYLGGLICPG------VLSSAG----ALSS 167
```

```
              ||    ||    :||  |      |        ||: ||      : || |    ::|
Db        419 LLYSDDAPGSLGNISVPSGATHVIFC--------GLV-PGAHYRVDIASSTGDISQSISG 469

Qy        168 RTAKLPRISLEVEEDSPVIGRSTTTSL 194
              |: ||  |||      || ||: : |
Db        470 YTSPLPPQSLE------VISRSSPSDL 490


RESULT 14
HXKA_YEAST
ID   HXKA_YEAST     STANDARD;      PRT;   485 AA.
AC   P04806;
DT   13-AUG-1987 (Rel. 05, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hexokinase A (EC 2.7.1.1) (Hexokinase PI).
GN   HXK1 OR HKA OR YFR053C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
X    NCBI_TaxID=4932;
     [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86120382; PubMed=3003701;
RA   Stachelek C., Stachelek J., Swan J., Botstein D., Konigsberg W.;
RT   "Identification, cloning and sequence determination of the genes
RT   specifying hexokinase A and B from yeast.";
RL   Nucleic Acids Res. 14:945-963(1986).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86083199; PubMed=3908224;
RA   Kopetzki E., Entian K.-D., Mecke D.;
RT   "Complete nucleotide sequence of the hexokinase PI gene (HXK1) of
RT   Saccharomyces cerevisiae.";
RL   Gene 39:95-102(1985).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=96287652; PubMed=8686379;
     Eki T., Naitou M., Hagiwara H., Ozawa M., Sasanuma S.-I.,
     Sasanuma M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
RT   "Analysis of a 36.2 kb DNA sequence including the right telomere of
RT   chromosome VI from Saccharomyces cerevisiae.";
RL   Yeast 12:149-167(1996).
RN   [5]
RP   ATP-BINDING, AND SEQUENCE OF 104-112.
RX   MEDLINE=88227998; PubMed=3131329;
RA   Tamura J.K., Ladime J.R., Cross R.L.;
RT   "The adenine nucleotide binding site on yeast hexokinase PII.
RT   Affinity labeling of Lys-111 by pyridoxal
RT   5'-diphospho-5'-adenosine.";
RL   J. Biol. Chem. 263:7907-7912(1988).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (3.5 ANGSTROMS).
RX   MEDLINE=81049624; PubMed=7001031;
RA   Bennett W.S. Jr., Steitz T.A.;
RT   "Structure of a complex between yeast hexokinase A and glucose. I.
RT   Structure determination and refinement at 3.5-A resolution.";
RL   J. Mol. Biol. 140:183-210(1980).
CC   -!- CATALYTIC ACTIVITY: ATP + D-hexose = ADP + D-hexose 6-phosphate.
CC   -!- ENZYME REGULATION: SUBJECT TO ALLOSTERIC CONTROL. SUBSTRATE
CC       INHIBITION BY ATP.
CC   -!- PATHWAY: FIRST STEP OF SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- MISCELLANEOUS: IN YEAST THERE ARE THREE GLUCOSE-PHOSPHORYLATING
CC       ISOENZYMES, DESIGNATED HEXOKINASE I, II AND GLUCOKINASE.
CC   -!- SIMILARITY: BELONGS TO THE HEXOKINASE FAMILY.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/H/HK.html".
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; M14410; AAA34698.1; -.
DR   EMBL; X03482; CAA27202.1; -.
DR   EMBL; D50617; BAA09292.1; -.
DR   PIR; A24531; KIBYHA.
DR   PIR; A28178; A28178.
DR   PDB; 1HKG; 15-OCT-91.
DR   SWISS-2DPAGE; P04806; YEAST.
DR   SGD; S0001949; HXK1.
DR   InterPro; IPR001312; Hexokinase.
DR   Pfam; PF00349; hexokinase; 1.
DR   Pfam; PF03727; hexokinase2; 1.
DR   PRINTS; PR00475; HEXOKINASE.
DR   ProDom; PD001109; Hexokinase; 1.
DR   PROSITE; PS00378; HEXOKINASES; 1.
KW   Transferase; Kinase; Glycolysis; Allosteric enzyme; ATP-binding;
KW   3D-structure.
FT   BINDING     111    111       ATP.
FT   DOMAIN      152    178       GLUCOSE-BINDING (POTENTIAL).
FT   CONFLICT     61     61       G -> V (IN REF. 1).
FT   CONFLICT    103    103       H -> R (IN REF. 1).
FT   CONFLICT    194    194       N -> K (IN REF. 1).
FT   CONFLICT    244    244       V -> C (IN REF. 1).
FT   CONFLICT    356    357       EN -> VF (IN REF. 2).
FT   CONFLICT    364    364       I -> M (IN REF. 1).
FT   CONFLICT    388    388       I -> T (IN REF. 2).
FT   CONFLICT    444    444       D -> EN (IN REF. 1).
FT   CONFLICT    479    480       SL -> VS (IN REF. 1).
FT   TURN         19     20
FT   HELIX        21     29
FT   HELIX        32     33
FT   HELIX        38     44
FT   TURN         45     46
FT   HELIX        47     54
FT   TURN         55     56
FT   STRAND       66     67
FT   STRAND       82     82
FT   STRAND       85     85
FT   STRAND       88     88
FT   TURN         89     90
FT   STRAND       91     92
FT   STRAND       94     94
FT   STRAND       97     97
FT   STRAND      111    111
FT   TURN        115    119
FT   STRAND      121    122
FT   TURN        124    126
FT   TURN        128    130
FT   HELIX       131    140
FT   HELIX       142    143
FT   STRAND      161    161
FT   STRAND      171    171
FT   TURN        176    177
FT   STRAND      180    181
FT   TURN        189    195
FT   HELIX       196    200
FT   TURN        211    216
FT   HELIX       218    222
FT   TURN        224    225
```

```
FT   STRAND      226    232
FT   STRAND      236    239
FT   STRAND      242    242
FT   TURN        244    245
FT   STRAND      264    264
FT   STRAND      266    267
FT   HELIX       270    272
FT   TURN        286    287
FT   HELIX       288    292
FT   TURN        293    294
FT   HELIX       300    303
FT   TURN        304    305
FT   TURN        308    310
FT   HELIX       311    323
FT   TURN        324    325
FT   TURN        339    340
FT   TURN        345    351
FT   HELIX       359    369
FT   TURN        375    377
FT   HELIX       378    396
FT   TURN        398    401
FT   HELIX       402    406
FT   TURN        407    408
FT   STRAND      412    417
FT   TURN        419    420
FT   HELIX       428    438
FT   STRAND      450    450
FT   STRAND      454    454
FT   TURN        458    460
FT   HELIX       461    474
FT   TURN        475    477
SQ   SEQUENCE   485 AA;  53738 MW;  AF5C9DA8F17BC3D0 CRC64;

  Query Match             7.2%;  Score 77.5;  DB 1;  Length 485;
  Best Local Similarity   21.5%;  Pred. No. 9.7;
  Matches   53;  Conservative   36;  Mismatches   93;  Indels   65;  Gaps   13;

Qy          4 HFLLFDIGNTNVKI------GIAVETAVLTSYVLPTDPGQT----------TDS-----I 42
              ::|  |:| ||:::      |        : | || |   |           ||     :
Db         81 NYLAIDLGGTNLRVVLVKLSGNHTFDTTQSKYKLPHDMRTTKHQEELWSFIADSLKDFMV 140

Qy         43 GLRLLEVLRHAGLG-----PADVGACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAI 97
                ||       ||     ||       : :  |:   ::|  : :    :    |   :
Db        141 EQELLNTKDTLPLGFTFSYPAS-----QNKINEGI---LQRWTKGF---DIPNVEGHDVV 189

Qy         98 PL----DNRYERPAEVGA------DRLVAAYAARRLYPGPRSLVSVDFGTATT---FDCV 144
              ||      :: | | |: |        |:|:|     |  | : : | |||      :| |
Db        190 PLLQNEISKRELPIEIVALINDTVGTLIASY-----YTDPETKMGVIFGTGVNGAFYDVV 244

Qy        145 EGGAYLGGLICPGVLSSA--------GALSSRTAKLPRISLE--VEEDSPVIGRSTTTSL 194
                  | | :   : |::        |:  :    |||   :  |:| ||  |:      :
Db        245 SDIEKLEGKLADDIPSNSPMAINCEYGSFDNEHLVLPRTKYDVAVDEQSPRPGQQAFEKM 304

Qy        195 NHGFIFG 201
                |:  |
Db        305 TSGYYLG 311


RESULT 15
RRPL_EBOSM
ID   RRPL_EBOSM     STANDARD;      PRT;  2210 AA.
AC   Q66802;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   RNA-directed RNA polymerase (EC 2.7.7.48) (Large structural protein)
DE   (L protein).
GN   L.
OS   Ebola virus (strain Sudan Maleo-79) (Ebo).
OC   Viruses; ssRNA negative-strand viruses; Mononegavirales; Filoviridae;
OC   Ebola-like viruses.
OX   NCBI_TaxID=128949;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sanchez A., Trappier S., Nichol S.T.;
RL   Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACTIVE POLYMERASE. IT
CC       MAY FUNCTION IN RNA SYNTHESIS, CAPPING, AS WELL AS METHYLATION OF
CC       CAPS, AND POLY(A) SYNTHESIS. ALSO INVOLVED IN TRANSCRIPTIONAL RNA
CC       EDITING.
CC   -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC       {RNA}(N).
CC   -!- SIMILARITY: PRIMARILY WITH THE N-TERMINAL HALF OF THE L PROTEINS
CC       OF RHABDOVIRUSES AND PARAMYXOVIRUSES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U23458; AAA79970.1; -.
DR   InterPro; IPR001016; Viral_RNA_pol_L.
DR   Pfam; PF00946; Paramyx_RNA_pol; 1.
KW   Transferase; RNA-directed RNA polymerase.
SQ   SEQUENCE   2210 AA;  251255 MW;  F50E6B624951AEA4 CRC64;

  Query Match             7.1%;  Score 77;  DB 1;  Length 2210;
  Best Local Similarity   23.3%;  Pred. No. 60;
  Matches   60;  Conservative   32;  Mismatches   82;  Indels   84;  Gaps   14;

Qy          9 DIGNTNVKIGIAVETAVL-TSYVLPTDPG---QTTDSIGLRLLEVLRHAGLG-------- 56
              |:  |   || | | ::  | ::||:       |  |:      :|:|  ||
Db        833 DLQGTLASIGTAFERSISETRHILPSRVAAAFHTYFSV-----RILQHHHLGFHKGSDLG 887

Qy         57 ------PADVGACVASSVVPGV-------NPLIRRACERYLYRKLLFAPGD-IAIPLDNR 102
                    | | |     |  || |       ||      |: ||| |    || :   |
Db        888 QLAINKPLDFGTIALSLAVPQVLGGLSFLNP------EKCLYRNL----GDPVTSGLFQL 937

Qy        103 YERPAEVGADRLVAAYAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPG----- 157
                : ||   :  |  |:      | :  ::||          |||  ||
Db        938 KHYLSMVGMSDIFHALVAK----SPGNCSAIDFVLNP-----------GGLNVPGSQDLT 982

Qy        158 -----VLSSAGALSSRTAKLPRI---SLEVEED---------SPVIGR------STTTSL 194
                  ::  :  ||:|  :  :   | ::|::          :||: |      | | |
Db        983 SFLRQIVRRSITLSARNKLINTLFHASADLEDELVCKWLLSSTPVMSRFAADIFSRTPSG 1042

Qy        195 NHGFIFGFAAMTEGVLAA 212
                | |:   |  :||:
Db       1043 KRLQILGYLEGTRTLLAS 1060


Search completed: June 24, 2003, 22:11:55
Job time : 5.26458 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 23.9973 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-59
Perfect score: 1077
Sequence: 1 MTQHFLLFDIGNTNVKIGIA.....SLNHGFIFGFAAMTEGVLAA 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1077	100.0	212	23	AAU91177	Pantothenate kinas
2	380.5	35.3	259	23	ABB47661	Listeria monocytog
3	379.5	35.2	262	23	AAU91170	Pantothenate kinas
4	352.5	32.7	258	23	AAU91172	Pantothenate kinas
5	348.5	32.4	219	23	AAU91176	Pantothenate kinas
6	338.5	31.4	233	23	AAU91163	Pantothenate kinas
7	338.5	31.4	255	23	AAU91154	Geobacter sulfurre
8	338.5	31.4	258	22	AAU01243	B. subtilis novel
9	338.5	31.4	258	23	AAU91149	Bacillus subtilis
10	330.5	30.7	254	23	AAU91171	Pantothenate kinas
11	327.5	30.4	250	23	AAU91150	Clostridium acetob
12	315.5	29.3	256	23	AAU91175	Pantothenate kinas
13	307.5	28.6	260	23	AAU91173	Pantothenate kinas
14	280	26.0	265	23	AAU91151	Streptomyces coeli
15	271	25.2	262	23	AAU91155	Deinococcus radiop
16	240.5	22.3	258	23	AAU91153	Rhodobacter capsul
17	221	20.5	246	23	AAU91156	Thermotoga maritim
18	214	19.9	257	23	AAU91174	Pantothenate kinas
19	209.5	19.5	272	22	AAG81225	Mycobacterium tube
20	209.5	19.5	272	23	AAU91152	Mycobacterium tube
21	192.5	17.9	273	23	AAU91157	Treponema pallidum
22	180.5	16.8	241	23	AAU91179	Pantothenate kinas
23	174	16.2	244	23	AAU91168	Pantothenate kinas
24	138	12.8	267	23	AAU91162	Bordella pertussis
25	136.5	12.7	592	20	AAY38615	Neisseria meningit
26	136.5	12.7	592	21	AAY74912	Neisseria meningit
27	136.5	12.7	592	23	AAU91169	Pantothenate kinas
28	134	12.4	249	23	AAU91182	Pantothenate kinas
29	133.5	12.4	455	20	AAY38617	Neisseria gonorrho
30	133.5	12.4	455	21	AAY74908	Neisseria gonorrhe
31	133.5	12.4	460	23	AAU91167	Pantothenate kinas
32	133.5	12.4	592	20	AAY38618	Neisseria gonorrho
33	133.5	12.4	592	21	AAY74911	Neisseria gonorrhe
34	132.5	12.3	455	21	AAY74910	Neisseria meningit
35	132.5	12.3	592	20	AAY38616	Neisseria meningit
36	132.5	12.3	592	21	AAY74913	Neisseria meningit
37	132.5	12.3	592	23	AAU91166	Pantothenate kinas
38	129	12.0	389	21	AAY74909	Neisseria meningit
39	123	11.4	189	20	AAY38614	Neisseria meningit
40	120.5	11.2	229	23	AAU91159	Aquifex aeolicus p
41	114.5	10.6	249	23	AAU91178	Pantothenate kinas
42	111	10.3	262	23	AAU91158	Borrelia burgdorfe
43	107	9.9	248	23	AAU91164	Pantothenate kinas
44	104	9.7	242	23	AAU91180	Pantothenate kinas
45	102	9.5	223	19	AAW98422	H. pylori GHPO 344

ALIGNMENTS

RESULT 1
AAU91177
ID AAU91177 standard; Protein; 212 AA.
XX
AC AAU91177;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #15.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Desulfovibrio vulgaris.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54198.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein

PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 107-108; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 212 AA;

Query Match 100.0%; Score 1077; DB 23; Length 212;
Best Local Similarity 100.0%; Pred. No. 6.5e-112;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60

Qy         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120

Qy        121 RRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 RRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVE 180

Qy        181 EDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212
              ||||||||||||||||||||||||||||||||
Db        181 EDSPVIGRSTTTSLNHGFIFGFAAMTEGVLAA 212
```

RESULT 2
ABB47661
ID ABB47661 standard; Protein; 259 AA.
XX
AC ABB47661;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes protein #365.
XX
KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease.
XX
OS Listeria monocytogenes.
XX
PN WO200177335-A2.
XX
PD 18-OCT-2001.
XX
PF 11-APR-2001; 2001WO-FR01118.
XX
PR 11-APR-2000; 2000FR-0004629.
XX
PA (INSP) INST PASTEUR.
XX
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX
DR WPI; 2002-010914/01.
XX
PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides -
XX
PS Claim 6; SEQ ID No 366; 192pp; French.
XX
CC The present invention relates to the genome sequence of Listeria
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in L.
CC monocytogenes and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 259 AA;

Query Match 35.3%; Score 380.5; DB 23; Length 259;
Best Local Similarity 39.4%; Pred. No. 6.2e-34;
Matches 82; Conservative 42; Mismatches 77; Indels 7; Gaps 5;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|: : :| : : || :|:| :|: :|    :| | |:|:   :
Db          3 LVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGIII 62

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :   | ||   | |:  || |   |: : : | |:|:||:| | ||   |
Db         63 SSVVPPIMHAMETMCVRYFNIRPLIVGPG-IKTGLNLKVDNPREIGSDRIVNAVAASEEY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
               |  :: |||||||||  : | | | || | ||:: |  || :| |||||:  :: | |
Db        122 GTP--VIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRV--DIAESS 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              :||:|| :|:  |  :||    ||:|
Db        178 QIIGKSTVSSMQAGIFYGFVGQCEGIIA 205
```

RESULT 3
AAU91170
ID AAU91170 standard; Protein; 262 AA.
XX
AC AAU91170;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #8.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus anthracis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX

```
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54191.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 8; Page 98-99; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   262 AA;

  Query Match             35.2%;  Score 379.5;  DB 23;  Length 262;
  Best Local Similarity   40.8%;  Pred. No. 8.1e-34;
  Matches   84;  Conservative   41;  Mismatches   74;  Indels    7;  Gaps    5;

Qy          7 LFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVAS 66
              : |:|||  :|:  |  :   : : ||  :| |  |: : ::| | ||   ||   : |
Db          4 VLDVGNTNAVLGVFEEGELRQHWRMETDRHKTEDEYGMLVKQLLEHEGLSFEDVKGIIVS 63

Qy         67 SVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||| :   : | ||:|   : |:  || |   |: :|| | ||||||:| | |   ||
Db         64 SVVPPIMFALERMCEKYFKIKPLVVGPG-IKTGLNIKYENPREVGADRIVNAVAGIHLYG 122

Qy        126 GPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
               |  |: ||||||||:  : |   |:||:| ||:: || || || ||||||  |: : |
Db        123 SP--LIIVDFGTATTYCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRI--EITKPSS 178

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:|::| :::  | ::|:    ||::
Db        179 VVGKNTVSAMQSGILYGYVGQVEGIV 204


RESULT 4
AAU91172
ID   AAU91172 standard; Protein; 258 AA.
XX
AC   AAU91172;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #10.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus stearothermophilus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54193.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 101-102; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   258 AA;

  Query Match             32.7%;  Score 352.5;  DB 23;  Length 258;
  Best Local Similarity   37.7%;  Pred. No. 8.3e-31;
  Matches   78;  Conservative   42;  Mismatches   80;  Indels    7;  Gaps    5;

Qy          7 LFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVAS 66
              : |:|||  :|:     :   : : |   :| |  |: :  :| | ||  :|:   : |
Db          4 VLDVGNTNTVLGVYDGDELKHHWRIETSRSKTEDEYGMMIKALLNHVGLQFSDIRGIIIS 63

Qy         67 SVVPGVNPLIRRACERYLYRK-LLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||| :   : | | :| : | |:  || |   || :|: | ||||||:| | |   ||
Db         64 SVVPPIMFALERMCLKYFHIKPLIVGPG-IKTGLDIKYDNPREVGADRIVNAVAGIHLYG 122

Qy        126 GPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
               |  |: ||||||||:  : |   |:|| | ||:: |  || :| ||||||  |:
Db        123 SP--LIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRI--EIARPDD 178

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              :||::| :::  | ::|:    ||:::
Db        179 IIGKNTVSAMQAGILYGYVGQVEGIVS 205


RESULT 5
AAU91176
ID   AAU91176 standard; Protein; 219 AA.
XX
AC   AAU91176;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #14.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Dehalococcoides ethenogenes.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
```

```
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54197.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 10; Page 106-107; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   219 AA;

  Query Match             32.4%;  Score 348.5;  DB 23;  Length 219;
  Best Local Similarity   38.5%;  Pred. No. 1.8e-30;
  Matches   80;  Conservative   39;  Mismatches   84;  Indels    5;  Gaps    3;

Qy          1 MTQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV 60
              |::  :  |||||:| |||     :| :: | :   :  |     || :|:|||: | ::
Db          1 MSEKLVAVDIGNTSVNIGIFEGEKLLANWHLGSVAQRMADEYASLLLGLLQHAGIHPEEL 60

Qy         61 GACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAA 120
                :  |||| :       : |     |     |   :  | : | ||||||:| | ||
Db         61 NRVIMCSVVPPLTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAAAA 120

Qy        121 RRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEV 179
              | ||||  : : || ||||||| : |||||:|| | ||: :|| |:: :|:|||:|  |:
Db        121 RVLYPG--ACIIVDMGTATTFDTLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKI--EI 176

Qy        180 EEDSPVIGRSTTTSLNHGFIFGFAAMTE 207
                 : ||| :| ::: | ||:  : |
Db        177 IRPAKVIGSNTVSAMQSGIYFGYIGLVE 204


RESULT 6
AAU91163
ID   AAU91163 standard; Protein; 233 AA.
XX
AC   AAU91163;
XX
DT   05-JUN-2002  (first entry)

DE   Pantothenate kinase (CoaX) #1.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus subtilis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54169.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Disclosure; Page 81-82; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   233 AA;

  Query Match             31.4%;  Score 338.5;  DB 23;  Length 233;
  Best Local Similarity   37.2%;  Pred. No. 2.6e-29;
  Matches   77;  Conservative   40;  Mismatches   83;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|:  :  :   : : |   :| |  |: |  :  |:||    :   :
Db          3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy         66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   : | | :| :    :  || :   |: :|: | |||||||:| | ||  ||
Db         63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  |: ||||||||:  : |   |:|| | ||:  |  || || ||||||  |:
Db        122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              :||::| :::  | :||:    ||::
Db        178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204


RESULT 7
AAU91154
ID   AAU91154 standard; Protein; 255 AA.
XX
AC   AAU91154;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Geobacter sulfurreducens pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Geobacter sulfurreducens.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
```

PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 72-73; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 255 AA;

Query Match 31.4%; Score 338.5; DB 23; Length 255;
Best Local Similarity 38.6%; Pred. No. 3e-29;
Matches 80; Conservative 38; Mismatches 82; Indels 7; Gaps 5;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:||||: :||     ::  : : ||  :|||  |: : |: | ||||   : | :
Db          3 LVIDVGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGILINELFRLAGLGLDQIRAVII 62

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :  :: |    |   | |:  || |   :  :|: | ||||||:| | |    |
Db         63 SSVVPPLTGVLERLSLGYFGMRPLVVGPG-IKTGMPIQYDNPREVGADRIVNAVAGYEKY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                ||: |||||||||| |   | | || | ||:: |  ||  | :||||:  ::   |
Db        122 --RTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPRV--DIIRPS 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
               :| |:|  |:  |  :|:  : : ::
Db        178 AIIARNTVNSMQAGIYYGYVGLVDEIV 204
```

RESULT 8
AAU01243
ID AAU01243 standard; Protein; 258 AA.
XX
AC AAU01243;
XX
DT 18-JUL-2001 (first entry)
XX
DE B. subtilis novel pantothenate kinase encoded by the gene coaX.
XX
KW Pantothenate kinase; coaX; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.
XX
OS Bacillus subtilis.
XX
PN WO200121772-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25993.
XX
PR 21-SEP-1999; 99US-0400494.
PR 07-JUN-2000; 2000US-0210072.
PR 28-JUL-2000; 2000US-0221836.
PR 24-AUG-2000; 2000US-0227860.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS.
XX
PI Yocum RR, Patterson TA, Hermann T, Pero JG;
XX
DR WPI; 2001-218644/22.
DR N-PSDB; AAS00984.
XX
PT New recombinant microorganism which overexpress a Bacillus subtilis
PT pantothenate biosynthetic enzyme, useful for the high yield production
PT of panto-compounds such as pantothenate and pantoate -
XX
PS Example 14; Fig 23; 292pp; English.
XX
CC The sequence represents a novel B. subtilis pantothenate kinase (encoded
CC by gene coaX), an enzyme of the pantothenate biosynthetic pathway.
CC Pantothenate, also known as vitamin B5, is used as a nutritional
CC supplement in mammals and humans. The invention concerns methods of
CC producing recombinant microorganisms overexpressing at least one Bacillus
CC subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
CC of producing them are useful for producing a panto-compound such as
CC pantothenate or pantoate, which is a nutritional requirement for
CC livestock and humans. The methods are also useful for the identification
CC of pantothenate kinase modulators. Panto-compounds are produced at a
CC significantly higher yield than prior art methods and can be produced
CC independent of the need to feed precursors which decreases expense.
XX
SQ Sequence 258 AA;

Query Match 31.4%; Score 338.5; DB 22; Length 258;
Best Local Similarity 37.2%; Pred. No. 3e-29;
Matches 77; Conservative 40; Mismatches 83; Indels 7; Gaps 5;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:||||  :|:  :  :   : : |   :| |  |: |  :  |:||    :   :
Db          3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy         66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              |||||  :   : | | :| :    :  || :   |: :|: | |||||||:| | ||  ||
Db         63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  |: ||||||||:  : |   |:|| | ||:  |  || || ||||||  |:
Db        122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
               :||::| :::  | :||:    ||::
Db        178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204
```

RESULT 9
AAU91149
ID AAU91149 standard; Protein; 258 AA.
XX
AC AAU91149;
XX
DT 05-JUN-2002 (first entry)
XX
DE Bacillus subtilis pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus subtilis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54168.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein

PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 67-68; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 258 AA;

Query Match 31.4%; Score 338.5; DB 23; Length 258;
Best Local Similarity 37.2%; Pred. No. 3e-29;
Matches 77; Conservative 40; Mismatches 83; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy 66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
Db 178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204

RESULT 10
AAU91171
ID AAU91171 standard; Protein; 254 AA.
XX
AC AAU91171;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #9.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bacillus halodurans.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54192.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 100; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 254 AA;

Query Match 30.7%; Score 330.5; DB 23; Length 254;
Best Local Similarity 36.7%; Pred. No. 2.3e-28;
Matches 76; Conservative 42; Mismatches 82; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGIVI 62

Qy 66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPMMFSLEQMCKKYFHVTPMIIGPG-IKTGLNIKYDNPKEVGADRIVNAVAAIELY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 GYP--AIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRI--EIAKPK 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
Db 178 QVVGTNTIDSMQSGIFYGYVSQVDGVV 204

RESULT 11
AAU91150
ID AAU91150 standard; Protein; 250 AA.
XX
AC AAU91150;
XX
DT 05-JUN-2002 (first entry)
XX
DE Clostridium acetobutylicum pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Clostridium acetobutylicum.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 68-69; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic

```
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   250 AA;

  Query Match            30.4%;  Score 327.5;  DB 23;  Length 250;
  Best Local Similarity  37.7%;  Pred. No. 4.9e-28;
  Matches   75;  Conservative   42;  Mismatches   75;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :||  :| :   : | ||  :: |  |:::: : :   | |  |   :
Db         20 LVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGVII 79

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :     :|     |:  || |   :: :|: | |||||:| | ||  :|
Db         80 SSVVPNIMYSLEHMIRKYFKINPLVVGPG-IKTGINIKYDNPKEVGADRIVNAVAAHEIY 138

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                 |||: :||||||||  | | | |||| ||||:  |: ||  : |||||:  |: : :
Db        139 --KRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRV--ELIKPA 194

Qy        184 PVIGRSTTTSLNHGFIFGF 202
                | ::| :|:  | :: :
Db        195 YAICKNTISSIQSGIVYRY 213


RESULT 12
AAU91175
ID   AAU91175 standard; Protein; 256 AA.
XX
AC   AAU91175;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #13.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Clostridium difficile.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54196.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 105; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   256 AA;

  Query Match            29.3%;  Score 315.5;  DB 23;  Length 256;
  Best Local Similarity  35.1%;  Pred. No. 1.1e-26;
  Matches   73;  Conservative   46;  Mismatches   82;  Indels    7;  Gaps    5;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |:||:|||| :||     ::  : : ||  :|:|  |: :  :  :  :  :|:   :
Db          3 LVFDVGNTNMVLGIYKGDKLVNYWRIKTDREKTSDEYGILISNLFDYDNVNISDIDDVII 62

Qy         66 SSVVPGVNPLIRRACERYLYRK-LLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| |   :   | :|  :: |:  || |   |: :|: | :|||||:| | |    |
Db         63 SSVVPNVMHSLENFCIKYCKKQPLIVGPG-IKTGLNIKYDNPKQVGADRIVNAVAGIEKY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                | |  |||||||||  : | | |||| | ||:  |: ||    :||||:  |: :
Db        122 GAPSIL--VDFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPRV--ELAKPG 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
                | :|| :::  | |:|:  : : :::
Db        178 MTICKSTVSAMQSGIIYGYVGLVDKIIS 205


RESULT 13
AAU91173
ID   AAU91173 standard; Protein; 260 AA.
XX
AC   AAU91173;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #11.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Caulobacter crescentus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54194.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 102-103; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
```

```
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   260 AA;

  Query Match             28.6%;  Score 307.5;  DB 23;  Length 260;
  Best Local Similarity   36.2%;  Pred. No. 9e-26;
  Matches   75;  Conservative   38;  Mismatches   91;  Indels    3;  Gaps    2;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |  : |||:    |    :  :   |:  :| |    : | ::|   |||   : | :
Db          3 LAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAVII 62

Qy         66 SSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||||     :|    ||   : |    :  : :| | |:|:| ||||||  |   |  :||
Db         63 SSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYP 122

Qy        126 GPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
              ||  || :| |||||||  |   ||: ||:| ||:  |  ||    ||||||:::    :
Db        123 GP--LVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              ::|  | :::  |  :|: :: ||::|
Db        181 IVGTDTVSAMQSGVFWGYISLIEGLVA 207


RESULT 14
AAU91151
ID   AAU91151 standard; Protein; 265 AA.
XX
AC   AAU91151;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Streptomyces coelicolor pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Streptomyces coelicolor.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.

     24-AUG-2000; 2000US-227860P.
     20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 69-70; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   265 AA;

  Query Match             26.0%;  Score 280;  DB 23;  Length 265;
  Best Local Similarity   35.5%;  Pred. No. 1.1e-22;
  Matches   76;  Conservative   36;  Mismatches   86;  Indels   16;  Gaps    8;

Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSI-----GLRLLEVLRHAGLGPADV 60
              |  |:|||:  :|:     ::  : : ||  :| | :     ||  :  |    ||
Db          3 LTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDGID 62

Qy         61 GACVASSVVPGVNPLIRRACERYL--YRKLLFAPG-DIAIPLDNRYERPAEVGADRLVAA 117
              |  : :: || |   :|    ||      :|  ||    :|:    : | |||||:: |
Db         63 GIAICAT-VPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPI--LTDHPKEVGADRIINA 119

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRIS 176
               ||  || ||   : |||||||||| |   | |:||:| ||:  |  ||  : |:| :|
Db        120 VAAVELYGGP--AIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKI- 176

Qy        177 LEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
               ||     |||::|  ::  | ::|||   :||:
Db        177 -EVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVV 209


RESULT 15
AAU91155
ID   AAU91155 standard; Protein; 262 AA.
XX
AC   AAU91155;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Deinococcus radiopugnans pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Deinococcus radiopugnans.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 73-74; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
```

SQ Sequence 262 AA;

Query Match 25.2%; Score 271; DB 23; Length 262;
Best Local Similarity 36.3%; Pred. No. 1.1e-21;
Matches 77; Conservative 32; Mismatches 87; Indels 16; Gaps 7;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLT-SYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV-GAC 63
              |  |||||   :|:|  :  || :: : |:     | : |:|  :   ||   | :  |
Db          7 LAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAG---APIPRAA 63

Qy         64 VASSVVPGVNPLIRRACERYLYRKLLFAPG----DIAIPLDNRYERPAEVGADRLVAAYA 119
              | ||| | |      | :|:              |: : ||     |  ||||||   :
Db         64 VLSSVAPPVGENYALALKRHFMIDAFAVSAENLPDVTVELDT----PGSVGADRLCNLFG 119

Qy        120 ARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLE 178
              | :  |    | |||||:| || |  |  :|||::  |   || || :| ||||||:|:
Db        120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy        179 VEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              | :  ||::|  :|  | :||:| | :|:|
Db        180 APETA--IGKNTVHALQSGLVFGYAEMVDGLL 209
```

Search completed: June 24, 2003, 22:08:39
Job time : 24.9973 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 8.74599 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-59
Perfect score: 1077
Sequence: 1 MTQHFLLFDIGNTNVKIGIA.....SLNHGFIFGFAAMTEGVLAA 212

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	382.5	35.5	259	2	AF1464	conserved hypothet
2	380.5	35.3	259	2	AF1102	conserved hypothet
3	338.5	31.4	233	2	S66100	conserved hypothet
4	335.5	31.2	273	2	E97293	probable transcrip
5	330.5	30.7	254	2	F83660	hypothetical prote
6	307.5	28.6	261	2	B87489	transcription acti
7	280	26.0	265	2	T36391	hypothetical prote
8	271	25.2	262	2	E75516	conserved hypothet
9	221	20.5	246	2	D72320	conserved hypothet
10	210.5	19.5	274	2	H86937	conserved hypothet
11	209.5	19.5	272	2	A70955	hypothetical prote
12	192.5	17.9	273	2	D71326	conserved hypothet
13	145.5	13.5	276	2	AI2292	hypothetical prote
14	138	12.8	267	2	I40327	baf protein - Bord
15	136.5	12.7	592	2	B81009	BirA protein/Bvg a
16	132.5	12.3	592	2	H82031	probable biotin-[a
17	120.5	11.2	229	2	E70465	hypothetical prote
18	111	10.3	262	2	F70165	conserved hypothet
19	107	9.9	248	2	H83111	hypothetical prote
20	104	9.7	242	2	A82637	conserved hypothet
21	102	9.5	223	2	F64627	hypothetical prote
22	97.5	9.1	699	2	A96529	hypothetical prote
23	97	9.0	223	2	G71887	hypothetical prote
24	97	9.0	400	2	F75275	chromate transport
25	96.5	9.0	308	2	F72508	probable hydrogena
26	94.5	8.8	577	2	H72739	probable molybdopt
27	92.5	8.6	257	2	S75559	hypothetical prote
28	91.5	8.5	312	2	S35222	acetaldehyde dehyd
29	91.5	8.5	458	2	F82372	potassium uptake p
30	89.5	8.3	312	2	T31270	acetaldehyde dehyd
31	89	8.3	380	2	D82088	chromate resistanc
32	89	8.3	387	2	T35425	probable fatty aci
33	87.5	8.1	668	2	S19739	integral membrane
34	87	8.1	349	2	H72714	probable O-sialogl
35	86.5	8.0	458	2	S50220	potassium transpor
36	86	8.0	458	2	AH0029	trk system potassi
37	85	7.9	209	2	H81382	hypothetical prote
38	84	7.8	236	2	AF0202	Mg(2+) transport A
39	84	7.8	392	2	G70594	probable MOEB- hom
40	83.5	7.8	452	2	G95356	probable oxidoredu
41	83	7.7	349	2	D75419	proline dipeptidas
42	83	7.7	474	2	C87556	hypothetical prote
43	83	7.7	641	2	PC7036	heat shock protein
44	82.5	7.7	392	2	B69411	3-ketoacyl-CoA thi
45	82.5	7.7	486	1	KIBYHB	hexokinase (EC 2.7

ALIGNMENTS

RESULT 1
AF1464
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip112
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1464
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1464
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0253
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 35.5%; Score 382.5; DB 2; Length 259;
Best Local Similarity 39.9%; Pred. No. 1e-27;
Matches 83; Conservative 40; Mismatches 78; Indels 7; Gaps 5;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:||||  :|:  |  :|  : : ||  :|:| :|: :|    :| | |:|:   :
Db          3 LVIDVGNTNCTVGVYKEQKLLRHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGIII 62

Qy         66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :   :   | ||  . | |:  || |   |: : : | |:|:||:| | ||   |
Db         63 SSVVPPIMHAMETMCVRYFNIRPLIVGPG-IKTGLNLKVDNPREIGSDRIVNAVAASEEY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
                |  :: |||||||||  : | | | || | ||:: |  || :| |||||:  :: | |
Db        122 GTP--VIVVDFGTATTFCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPRV--DIAESS 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
                :||:||  |:  |  :||    ||::|
Db        178 QIIGKSTVASMQAGIFYGFIGQCEGIIA 205
```

RESULT 2
AF1102
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain E
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1102
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1102
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0221
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 35.3%; Score 380.5; DB 2; Length 259;
Best Local Similarity 39.4%; Pred. No. 1.6e-27;
Matches 82; Conservative 42; Mismatches 77; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGIII 62

Qy 66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPIMHAMETMCVRYFNIRPLIVGPG-IKTGLNLKVDNPREIGSDRIVNAVAASEEY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 GTP--VIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRV--DIAESS 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
Db 178 QIIGKSTVSSMQAGIFYGFVGQCEGIIA 205

RESULT 3
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S66100; E69740
R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis chrom
A;Reference number: S65967; MUID:96051385; PMID:7584024
A;Accession: S66100
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <OGA>
A;Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g467459
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1993
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:el182(
A;Experimental source: strain 168
C;Genetics:
A;Gene: yacB
A;Start codon: TTG
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 31.4%; Score 338.5; DB 2; Length 233;
Best Local Similarity 37.2%; Pred. No. 1e-23;
Matches 77; Conservative 40; Mismatches 83; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62

Qy 66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177

Qy 184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
Db 178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204

RESULT 4
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostr
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; I
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 31.2%; Score 335.5; DB 2; Length 273;
Best Local Similarity 38.2%; Pred. No. 2.4e-23;
Matches 76; Conservative 42; Mismatches 74; Indels 7; Gaps 5;

Qy 6 LLFDIGNTNVKIGIAVETAVLTSYVLPTPDGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
Db 14 LVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGVII 73

Qy 66 SSVVPGVNPLIRRACERYL-YRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
Db 74 SSVVPNIMYSLEHMIRKYFKINPLVVGPG-IKTGINIKYDNPKEVGADRIVNAVAAHEIY 132

Qy 125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
Db 133 --KRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRV--ELIKPA 188

Qy 184 PVIGRSTTTSLNHGFIFGF 202
Db 189 YAICKNTISSIQSGIVYGY 207

RESULT 5
F83660

hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83660
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03805.1; GSPDB:GN0
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0086
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 30.7%; Score 330.5; DB 2; Length 254;
Best Local Similarity 36.7%; Pred. No. 6.3e-23;
Matches 76; Conservative 42; Mismatches 82; Indels 7; Gaps 5;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |: |:|||| :|: : :: : | |  :| |  : : :  |||| |: |
Db          3 LVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGIVI 62

Qy         66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124
              ||||| :  : : |::| :  ::  || |  |: :|: | ||||||:| | ||  ||
Db         63 SSVVPPMMFSLEQMCKKYFHVTPMIIGPG-IKTGLNIKYDNPKEVGADRIVNAVAAIELY 121

Qy        125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183
               |  : ||||||||:  : |  | ||:| ||:: |  ||  | :|||||  |: :
Db        122 GYP--AIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRI--EIAKPK 177

Qy        184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:| :|  |:  |  :|: :  :||:
Db        178 QVVGTNTIDSMQSGIFYGYVSQVDGVV 204
```

RESULT 6
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
roc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 28.6%; Score 307.5; DB 2; Length 261;
Best Local Similarity 36.2%; Pred. No. 8.5e-21;
Matches 75; Conservative 38; Mismatches 91; Indels 3; Gaps 2;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65
              |  : ||||    |    : :  :   |:  :| |   : | ::|   |||   : | :
Db          4 LAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAVII 63

Qy         66 SSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLYP 125
              |||||     :|    ||   : |    :  : :| | |:|:| |||||| |  |  :||
Db         64 SSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMVYP 123

Qy        126 GPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDSP 184
```

```
              ||  || :| ||||||| |   ||: ||:| ||:  |  ||    ||||||:::    :
Db        124 GP--LVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 181

Qy        185 VIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              ::|  | :::  |  :|: :: ||::|
Db        182 IVGTDTVSAMQSGVFWGYISLIEGLVA 208
```

RESULT 7
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE94.31c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 26.0%; Score 280; DB 2; Length 265;
Best Local Similarity 35.5%; Pred. No. 3e-18;
Matches 76; Conservative 36; Mismatches 86; Indels 16; Gaps 8;

```
Qy          6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSI-----GLRLLEVLRHAGLGPADV 60
              |  |:|||:  :|:     ::  : : ||  :| | :     ||  :  |    ||
Db          3 LTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDGID 62

Qy         61 GACVASSVVPGVNPLIRRACERYL--YRKLLFAPG-DIAIPLDNRYERPAEVGADRLVAA 117
              |  : :: || |   :|    ||      :|  ||    :|:    : | ||||||: |
Db         63 GIAICAT-VPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPI--LTDHPKEVGADRIINA 119

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRIS 176
              ||  || ||   : ||||||||||| |   | |:||:| ||:  |  ||  : |:| :|
Db        120 VAAVELYGGP--AIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKI- 176

Qy        177 LEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              ||     |||::|  ::  | ::|||   :||:
Db        177 -EVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVV 209
```

RESULT 8
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: E75516
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <WHI>
A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0461
A;Map position: 1
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 25.2%; Score 271; DB 2; Length 262;
Best Local Similarity 36.3%; Pred. No. 2e-17;

```
  Matches   77;  Conservative   32;  Mismatches   87;  Indels   16;  Gaps    7;

Qy          6 LLFDIGNTNVKIGIAVETAVLT-SYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADV-GAC 63
              |  |||||   :|:|  :  || :: : |:     | : |:|  :   ||   | :  |
Db          7 LAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAG---APIPRAA 63

Qy         64 VASSVVPGVNPLIRRACERYLYRKLLFAPG----DIAIPLDNRYERPAEVGADRLVAAYA 119
              | ||| | |       | :|:               |: : ||     |  ||||||   :
Db         64 VLSSVAPPVGENYALALKRHFMIDAFAVSAENLPDVTVELDT----PGSVGADRLCNLFG 119

Qy        120 ARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLE 178
              | :   |    | |||||:| || |  |  :|||::  |   || || :| ||||||:|:
Db        120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy        179 VEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                | :  ||::|  :|  | :||:| | :|:|
Db        180 APETA--IGKNTVHALQSGLVFGYAEMVDGLL 209


RESULT 9
D72320
nserved hypothetical protein - Thermotoga maritima (strain MSB8)
Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498141
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             20.5%;  Score 221;  DB 2;  Length 246;
  Best Local Similarity   31.9%;  Pred. No. 7.3e-13;
  Matches   68;  Conservative   31;  Mismatches   92;  Indels   22;  Gaps    7;

Qy          5 FLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPA--DVGA 62
              :|| |:|||:    |  :      : | |   || | :   |     |  || |  ::
Db          2 YLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDELFSHL-----HPLLGDAMREIKG 56

            63 CVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDN-----RYERPAEVGADRLVAA 117
               :|||  | :| |  ::|      |    | :  |        : |:||||||:
Db         57 IGVASVVPTQNTVIERFSQKY------FHISPIWVKAKNGCVKWNVKNPSEVGADRVANV 110

Qy        118 YAARRLYPGPRSLVSVDFGTATTFDCVEGGAYLGGLICPGVLSSAGALSSRTAKLPRISL 177
              |  : | |   :: :| ||||| | |  |:| || | ||      :|   ||||| :
Db        111 VAFVKEY-GKNGII-IDMGTATTVDLVVNGSYEGGAILPGFFMMVHSLFRGTAKLPLV-- 166

Qy        178 EVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              ||:    |:|: |  ::  | : |     ||::
Db        167 EVKPADFVVGKDTEENIRLGVVNGSVYALEGII 199


RESULT 10
H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             19.5%;  Score 210.5;  DB 2;  Length 274;
  Best Local Similarity   31.2%;  Pred. No. 7.7e-12;
  Matches   67;  Conservative   41;  Mismatches   86;  Indels   21;  Gaps   10;

Qy          6 LLFDIGNTNVKIGI----AVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              |  |: ||:  :|:         |:  : : |:   | | : | ::: |   |      :
Db          3 LAIDVRNTHTVVGLLSGSKEHAKVVQQWRIRTESEVTADELAL-IIDGL--IGDDSERLA 59

Qy         62 ACVASSVVPGVNPLIRRACERY--LYRKLLFAPG---DIAIPLDNRYERPAEVGADRLVA 116
                | | || |  :|   ::|      :|  ||    | : :||    | |||||:|
Db         60 GAAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDN----PKEVGADRIVN 115

Qy        117 AYAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKLPRI 175
                ||   : |  ::| ||||::   | |   | :||| | |||  |: | ::|:| | |:
Db        116 CLAAFHKF-GQAAIV-VDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRV 173

Qy        176 SLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:     |:|::|   :  | :|||| : :|::
Db        174 --ELARPRSVVGKNTVECMQAGVVFGFAGLVDGLV 206


RESULT 11
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70955
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genc
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70955
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-272 <COL>
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g21139
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3600c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             19.5%;  Score 209.5;  DB 2;  Length 272;
  Best Local Similarity   30.3%;  Pred. No. 9.5e-12;
  Matches   66;  Conservative   41;  Mismatches   84;  Indels   27;  Gaps    9;

Qy          6 LLFDIGNTNVKIGI----AVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              |  |: ||:  :|:         |:  : : |:   | | : | :  ::   |      :
Db          3 LAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLI---GEDSERLT 59

Qy         62 ACVASSVVPGVNPLIRRACERY--LYRKLLFAPG---DIAIPLDNRYERPAEVGADRLV- 115
                | | || |   :|   ::|      :|  ||    | : :||    | |||||:|
Db         60 GTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDN----PKEVGADRIVN 115

Qy        116 --AAYAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAKL 172
                 |||   |     :: : ||||::   | |   | :||| | |||  |: | ::|:| |
Db        116 CLAAYDRFR-----KAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAAL 170

Qy        173 PRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
                |:  |:     |:|::|   :  | :|||| : :|::
```

```
Db        171 RRV--ELARPRSVVGKNTVECMQAGAVFGFAGLVDGLV 206

RESULT 12
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C;Accession: D71326
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71326
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <COL>
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0431
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             17.9%;  Score 192.5;  DB 2;  Length 273;
  Best Local Similarity   30.9%;  Pred. No. 3.5e-10;
  Matches   68;  Conservative   35;  Mismatches   92;  Indels   25;  Gaps    8;

Qy          6 LLFDIGNTNVKIGIAVE----TAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              || |:||::|  ||  |       |   : |  |  :| |   | :  :   ||:| | :
Db          2 LLIDVGNSHVVFGIQGENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASLR 61

Qy         62 ACVASSVVPGVNPLIRRACERYL-YRKLLFAPG-----DIAIPLDNRYERPAEVGAD--- 112
                |||||  :   |  |  :    : ::| |        : ||   |    ||:| |
Db         62 DAFISSVVPVLTKTIADAVAQISGVQPVVFGPWAYEHLPVRIPEPVR----AEIGTDLVA 117

Qy        113 RLVAAYAARRLYPGPRSLVSVDFGTATTFDCVEG-GAYLGGLICPGVLSSAGALSSRTAK 171
              ||||   |      : | || ||| ||  |:| |   |  | ||: ::  :| : ||:
Db        118 NAVAAYVHFR-----SACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQ 172

Qy        172 LPRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVLA 211
              || : | : :   |:|: || ::  | : |   :   ::|
Db        173 LPLVPLALPDS--VLGKDTTHAVQAGVVRGTLFVIRAMIA 210


RESULT 13
I2292
ypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

  Query Match             13.5%;  Score 145.5;  DB 2;  Length 276;
  Best Local Similarity   27.4%;  Pred. No. 7.6e-06;
  Matches   58;  Conservative   35;  Mismatches   96;  Indels   23;  Gaps    9;

Qy          5 FLLFDIGNTNVK----IGIAVETAVLTSYVLPTD------PGQTTDSIGLRLLEVLRHAG 54
              :|  :|||: :     :| ::|    | | ||        |:|   :|    |:

Db         14 WLALEIGNSRLHWALFMGESLEFTWDTEY-LPESVIQQLGNGETKLEVGSEEKEIF--FT 70

Qy         55 LGPADVGACVASSVVPGVNPLIRRACERYLYRKLLFAPGDIAIPLDNRYERPAEVGADRL 114
                |      |      :  | |       | ||  :::        |||:| |     :| ||
Db         71 FFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLD---QIPLNNIY---PTLGIDRA 124

Qy        115 VAAYAARRLYPGPRSLVSVDFGTATTFDCVEGGAYL-GGLICPGVLSSAGALSSRTAKLP 173
              :| : |   :  |  :: :| ||| ||   :||  | || | |||     :|  :| :||
Db        125 LALWGAGMSWGFP--VLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQLP 182

Qy        174 RISLE-VEEDSPVIGRSTTTSLNHGFIFGFAA 204
              :: :| ::   |    :|| ::  | |:   |
Db        183 QVEMEAIKSLPPRFALNTTEAIQSGVIYTLIA 214


RESULT 14
I40327
baf protein - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40327; S70669
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
J. Bacteriol. 177, 3801-3807, 1995
A;Title: Identification of a Bordetella pertussis regulatory factor required for tra
A;Reference number: I40327; MUID:95325323; PMID:7601846
A;Accession: I40327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229
R;Allen, A.; Maskell, D.
Mol. Microbiol. 19, 37-52, 1996
A;Title: The identification, cloning and mutagenesis of a genetic locus required for
A;Reference number: S70669; MUID:96419162; PMID:8821935
A;Accession: S70669
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 239-267 <ALL>
A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: baf

  Query Match             12.8%;  Score 138;  DB 2;  Length 267;
  Best Local Similarity   27.4%;  Pred. No. 3.6e-05;
  Matches   62;  Conservative   34;  Mismatches   78;  Indels   52;  Gaps   12;

Qy          6 LLFDIGNTNVKIG--------------IAVETAVLTSY--VLPTDPGQTTDSIGLRLLE 48
              :| | ||: :|:|              :| :   | :    | | | :   ::|: :
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNV-- 60

Qy         49 VLRHAGLGPADVGACVASSVVPGVNPLIRRACE-RYLYRKLLFAPGDIAIPLDNRYERPA 107
                |||      |  :|:::  |        |: |:|  :       :|: | | |  |
Db         61 ----AGLAR---GEAIAATLRAG-------GCDIRWLRAQ------PLAMGLRNGYRNPD 100

Qy        108 EVGADR--LVAAYAARRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGA 164
              ::||||   :    ||:    |  ||:  |||||| | :     : |||| ||     ||
Db        101 QLGADRWACMVGVLARQPSVHPPLLVA-SFGTATTLDTIGPDNVFPGGLILPGPAMMRGA 159

Qy        165 LSSRTAKLPRISLEVEEDSPVIGRSTTTSLNHGFIFGFAAMTEGVL 210
              |:  || || ::  :  | |:       |       | ||   | :
Db        160 LAYGTAHLP-LADGLVADYPIDTHQAIAS-------GIAAAQAGAI 197


RESULT 15
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strai
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
```

Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

Query Match 12.7%; Score 136.5; DB 2; Length 592;
Best Local Similarity 30.9%; Pred. No. 0.00013;
Matches 54; Conservative 18; Mismatches 86; Indels 17; Gaps 6;

```
Qy          2 TQHFLLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVG 61
              :: ||| | ||: :|    ||                |  |   | |  |       ||
Db        337 SERFLLLDGGNSRLK-WAWVENGTFA----------TVGSAPYRDLSPLGAEWAEKADGN 385

Qy         62 ACVASSVVPGVNPLIRRACERYLYRKLLFAPGDI-AIPLDNRYERPAEVGADRLVAAYAA 120
                :    | |     :   :  | ||: : |    |: : | |  | | |:||   |  :
Db        386 VRIVGCAVCG--EFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGS 443

Qy        121 RRLYPGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPR 174
              ||      : | |  ||| | | : : | |||| | ||      :|: ||| | |
Db        444 RRF--SRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 496
```

Search completed: June 24, 2003, 22:19:22
Job time : 9.74599 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:49:47 ; Search time 9.81862 Seconds
(without alignments)
725.188 Million cell updates/sec

Title: US-09-813-453A-65
Perfect score: 1230
Sequence: 1 MNDWLFDLGNSRFKCASLRE.....HRPTLVLDGLAIWAAVAANV 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	89	7.2	446	6	5310667-4	Patent No. 5310667
2	88.5	7.2	3491	2	US-07-642-734C-2	Sequence 2, Appli
3	88.5	7.2	3491	3	US-08-439-009A-2	Sequence 2, Appli
4	87	7.1	446	4	US-09-243-374-4	Sequence 4, Appli
5	83.5	6.8	5588	4	US-09-036-987A-6	Sequence 6, Appli
6	83.5	6.8	5588	4	US-09-370-700-6	Sequence 6, Appli
7	83	6.7	730	2	US-08-696-944-2	Sequence 2, Appli
8	81.5	6.6	654	1	US-08-441-139-11	Sequence 11, Appl
9	80.5	6.5	666	1	US-08-441-139-16	Sequence 16, Appl
10	80	6.5	469	4	US-09-044-718-82	Sequence 82, Appl
11	80	6.5	989	4	US-09-199-637A-273	Sequence 273, App
12	79	6.4	829	4	US-09-413-814-105	Sequence 105, App
13	79	6.4	2475	4	US-09-413-814-48	Sequence 48, Appl
14	78.5	6.4	510	3	US-08-801-344-8	Sequence 8, Appli
15	78.5	6.4	510	4	US-09-498-599-8	Sequence 8, Appli
16	78.5	6.4	2183	1	US-08-348-891A-7	Sequence 7, Appli
17	78.5	6.4	2183	2	US-08-905-817-7	Sequence 7, Appli
18	78.5	6.4	3816	4	US-09-428-517-3	Sequence 3, Appli
19	78.5	6.4	6095	4	US-09-144-085-2	Sequence 2, Appli
20	78	6.3	4150	4	US-09-428-517-2	Sequence 2, Appli
21	77.5	6.3	819	4	US-09-651-656-15	Sequence 15, Appl
22	77.5	6.3	819	4	US-09-650-855-15	Sequence 15, Appl
23	77	6.3	1480	3	US-09-191-647-7	Sequence 7, Appli
24	77	6.3	1480	4	US-09-540-245A-7	Sequence 7, Appli
25	77	6.3	1480	4	US-09-540-153-7	Sequence 7, Appli
26	77	6.3	1480	4	US-09-182-024A-5	Sequence 5, Appli
27	77	6.3	1480	5	PCT-US91-09055-2	Sequence 2, Appli
28	77	6.3	5087	4	US-09-144-085-1	Sequence 1, Appli
29	77	6.3	15281	2	US-08-471-119A-2	Sequence 2, Appli
30	76.5	6.2	334	4	US-09-463-702A-38	Sequence 38, Appl
31	76.5	6.2	370	2	US-08-559-303B-77	Sequence 77, Appl
32	76.5	6.2	370	4	US-09-175-828-77	Sequence 77, Appl
33	76.5	6.2	607	3	US-08-781-891-75	Sequence 75, Appl
34	76	6.2	422	5	PCT-US91-01360-2	Sequence 2, Appli
35	75.5	6.1	106	4	US-09-249-542-20	Sequence 20, Appl
36	75.5	6.1	249	4	US-09-372-422A-30	Sequence 30, Appl
37	75.5	6.1	497	1	US-08-513-841-2	Sequence 2, Appli
38	75.5	6.1	497	2	US-08-696-834-2	Sequence 2, Appli
39	75.5	6.1	497	2	US-08-942-673-2	Sequence 2, Appli
40	75.5	6.1	497	4	US-09-118-317-2	Sequence 2, Appli
41	75.5	6.1	601	4	US-09-155-036-5	Sequence 5, Appli
42	75	6.1	355	2	US-08-758-621-4	Sequence 4, Appli
43	75	6.1	355	4	US-09-107-858-4	Sequence 4, Appli
44	75	6.1	1022	3	US-08-772-270A-2	Sequence 2, Appli
45	75	6.1	3798	3	US-09-335-409-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
5310667-4
;Patent No. 5310667
; APPLICANT: EICHHOLTZ, DAVID A.;GASSER, CHARLES S.;KISHORE,
;GANESH M.
; TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL
;-3-PHOSPHOSHIKIMATE SYNTHASES
; NUMBER OF SEQUENCES: 37
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/380,963
; FILING DATE: 17-JUL-1989
;SEQ ID NO:4:
; LENGTH: 446
5310667-4

Query Match 7.2%; Score 89; DB 6; Length 446;
Best Local Similarity 24.8%; Pred. No. 0.078;
Matches 64; Conservative 27; Mismatches 97; Indels 70; Gaps 12;

```
Qy         18 LREGVIGP-VTVLPYLTETMDAF----------ALQELPRGRVAYLASVAAPAITTHVL 65
              :||  ||  |  |  |   :| |              || |:|    ||::  :|  ::
Db        131 MRERPIGDLVAGLKQLNADVDCFLGTNCPPVRVNGKGGLPGGKVKLSGSVSSQYLTALLM 190

Qy         66 EV-LKIHFEQVQVAATVAACAGVRIAYAHPERFGV--------DRFLALLGS--YGEGNV 114
                 | :   ::::   : :   | :     |||||        ||||   |      ||
Db        191 AAPLALGDVEIEIVDKLISVPYVEMTLKLMERFGVSVEHSGNWDRFLVHGGQKYKSPGNA 250

Qy        115 LVVG---------VGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYL 165
                | |          | |:|   :  |||   |  |                  | |   :
Db        251 FVEGDASSASYLLAGAAITGGTITVNGC---GTSS----------------LQGD--V 287

Qy        166 EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHG--------GGVASLLPW 217
              :|||  |        || |   | |:  :    | | | :| |          ||  |
Db        288 KFAEVLEKM------GAKVTWSENSVTVSGPPRDFSGRKVLRGIDVNMNKMPDVAMTLAV 341

Qy        218 LGDVVHRPTLVLDGLAIW 235
              :   : || : | :| |
Db        342 VALFANGPTAIRD-VASW 358
```

RESULT 2
US-07-642-734C-2
; Sequence 2, Application US/07642734C
; Patent No. 5824513
; GENERAL INFORMATION:
; APPLICANT: Katz, L
; APPLICANT: Donadio, S
; APPLICANT: Mcalpine, J B
; TITLE OF INVENTION: Recombinant DNA Method for Producing
; TITLE OF INVENTION: Erythromycin Analogs

```
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3491 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-642-734C-2

  Query Match             7.2%;  Score 88.5;  DB 2;  Length 3491;
  Best Local Similarity   23.1%;  Pred. No. 2.3;
  Matches   63;  Conservative   29;  Mismatches  102;  Indels   79;  Gaps   12;

Qy         12 RFKCASLREGV-IGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVL----- 65
              | :  | |  | || | |    :| :|  |      |   : :  || |:     :
Db          5 RSRTTSRRTPVRIGAVVVASSTSELLDGLAAVADGRPHASVVRGVARPSAPVVFVFPGQG 64

Qy         66 --------EVLKIHFEQVQVAATVAACAGV-------------------RIAYAHPERF 97
                    |:|    |    || : |||                         |:    |  |
Db         65 AQWAGMAGELLG---ESRVFAAAMDACARAFEPVTDWTLAQVLDSPEQSRRVEVVQPALF 121

Qy         98 GVDRFLALL-GSYGEGNVLVVG--VGTALTIDLLAANGCHLGGRISASPTLMRQALHARA 154
               |   || |  |:|     |||  :|     :| ||: |   |   |:       || :|
Db        122 AVQTSLAALWRSFGVTPDAVVGHSIG-----ELAAAHVC---GAAGAADAARAAALWSR- 172

Qy        155 EQLPLSGGNYL------------EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQ-- 200
              | :|| |   :               |   :| :::| ||    |:  |     :|: :
Db        173 EMIPLVGNGDMAAVALSADEIEPRIARWDDDVVLAGVNGPRSVLLTGSPEPVARRVQELS 232

Qy        201 ---------SVRLLLHGG-------GVASLLPW 217
                     :| :  |          |: | | |
Db        233 AEGVRAQVINVSMAAHSAQVDDIAEGMRSALAW 265


RESULT 3
US-08-439-009A-2
; Sequence 2, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3491 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-439-009A-2

  Query Match             7.2%;  Score 88.5;  DB 3;  Length 3491;
  Best Local Similarity   23.1%;  Pred. No. 2.3;
  Matches   63;  Conservative   29;  Mismatches  102;  Indels   79;  Gaps   12;

Qy         12 RFKCASLREGV-IGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVL----- 65
              | :  | |  | || | |    :| :|  |      |   : :  || |:     :
Db          5 RSRTTSRRTPVRIGAVVVASSTSELLDGLAAVADGRPHASVVRGVARPSAPVVFVFPGQG 64

Qy         66 --------EVLKIHFEQVQVAATVAACAGV-------------------RIAYAHPERF 97
                    |:|    |    || : |||                         |:    |  |
Db         65 AQWAGMAGELLG---ESRVFAAAMDACARAFEPVTDWTLAQVLDSPEQSRRVEVVQPALF 121

Qy         98 GVDRFLALL-GSYGEGNVLVVG--VGTALTIDLLAANGCHLGGRISASPTLMRQALHARA 154
               |   || |  |:|     |||  :|     :| ||: |   |   |:       || :|
Db        122 AVQTSLAALWRSFGVTPDAVVGHSIG-----ELAAAHVC---GAAGAADAARAAALWSR- 172

Qy        155 EQLPLSGGNYL------------EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQ-- 200
              | :|| |   :               |   :| :::| ||    |:  |     :|: :
Db        173 EMIPLVGNGDMAAVALSADEIEPRIARWDDDVVLAGVNGPRSVLLTGSPEPVARRVQELS 232

Qy        201 ---------SVRLLLHGG-------GVASLLPW 217
                     :| :  |          |: | | |
Db        233 AEGVRAQVINVSMAAHSAQVDDIAEGMRSALAW 265


RESULT 4
US-09-243-374-4
; Sequence 4, Application US/09243374B
; Patent No. 6225114
; GENERAL INFORMATION:
;  APPLICANT: Eichholtz, David Alan
;  APPLICANT: Gasser, Charles Scott
;  APPLICANT: Kishore, Ganesh M
;  TITLE OF INVENTION: Glyphosate-tolerant 5-enolpyruvyl-3-phosphoshikimate
;  TITLE OF INVENTION: synthases
;  FILE REFERENCE: EPSP synthases
;  CURRENT APPLICATION NUMBER: US/09/243,374B
;  CURRENT FILING DATE: 1999-02-01
;  EARLIER APPLICATION NUMBER: 07/590,647
;  EARLIER FILING DATE: 1990-09-28
;  NUMBER OF SEQ ID NOS: 34
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
;   LENGTH: 446
;   TYPE: PRT
;   ORGANISM: Glycine max
```

US-09-243-374-4

Query Match 7.1%; Score 87; DB 4; Length 446;
Best Local Similarity 24.8%; Pred. No. 0.13;
Matches 64; Conservative 27; Mismatches 97; Indels 70; Gaps 12;

```
Qy         18 LREGVIGP-VTVLPYLTETMDAF----------ALQELPRGRVAYLASVAAPAITTHVL 65
Db        131 MRERPIGDLVAGLKQLGADVDCFLGTNCPPVRVNGKGGLPGGKVKLSGSVSSQYLTALLM 190

Qy         66 EV-LKIHFEQVQVAATVAACAGVRIAYAHPERFGV--------DRFLALLGS--YGEGNV 114
Db        191 AAPLALGDVEIEIVDKLISVPYVEMTLKLMERFGVSVEHSGNWDRFLVHGGQKYKSPGNA 250

Qy        115 LVVG---------VGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYL 165
Db        251 FVEGDASSASYLLAGAAITGGTITVNGC---GTSS----------------LQGD--V 287

Qy        166 EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHG--------GGVASLLPW 217
Db        288 KFAEVLEKM------GAKVTWSENSVTVSGPPRDFSGRKVLRGIDVNMNKMPDVAMTLAV 341

Qy        218 LGDVVHRPTLVLDGLAIW 235
Db        342 VALFANGPTAIRD-VASW 358
```

RESULT 5
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-6

Query Match 6.8%; Score 83.5; DB 4; Length 5588;
Best Local Similarity 28.3%; Pred. No. 19;
Matches 54; Conservative 15; Mismatches 63; Indels 59; Gaps 10;

```
Qy         40 ALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAAT---------VAACAG---- 86
Db       4807 AMQELP---------TAARRATQQAMAMLQQWLADDRFVSTRLILLTHRAVSAVAGEDVL 4857

Qy         87 ----------VRIAYA-HPERFGVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHL 135
Db       4858 DLVHAPLWGLVRSAQAEHPDRF------ALIDMDDER-----ASQTALA-EALTAGEAQL 4905

Qy        136 GGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAH 195
Db       4906 AVRSGV-------VLAPRLGQVKVSGGEAFRWDEGT--VLVTGGTGGLGALLARHLVSAH 4956

Qy        196 QRLDQSVRLLL 206
Db       4957 -----GVRHLL 4962
```

RESULT 6
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5588
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

Query Match 6.8%; Score 83.5; DB 4; Length 5588;
Best Local Similarity 28.3%; Pred. No. 19;
Matches 54; Conservative 15; Mismatches 63; Indels 59; Gaps 10;

```
Qy         40 ALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAAT---------VAACAG---- 86
Db       4807 AMQELP---------TAARRATQQAMAMLQQWLADDRFVSTRLILLTHRAVSAVAGEDVL 4857

Qy         87 ----------VRIAYA-HPERFGVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHL 135
Db       4858 DLVHAPLWGLVRSAQAEHPDRF------ALIDMDDER-----ASQTALA-EALTAGEAQL 4905

Qy        136 GGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAH 195
Db       4906 AVRSGV-------VLAPRLGQVKVSGGEAFRWDEGT--VLVTGGTGGLGALLARHLVSAH 4956

Qy        196 QRLDQSVRLLL 206
Db       4957 -----GVRHLL 4962
```

RESULT 7
US-08-696-944-2
; Sequence 2, Application US/08696944
; Patent No. 5981831

```
;  GENERAL INFORMATION:
;    APPLICANT:  Sumant CHENGAPPA
;    APPLICANT:  Susan A. HELLYER
;    APPLICANT:  John S. REID
;    APPLICANT:  Jacqueline DE SILVA
;    TITLE OF INVENTION:  No. 5981831el Exo-(1-4)-Beta-D Galactanase
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pillsbury Madison & Sutro, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  MS Word
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,944
;      FILING DATE:  23-AUG-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB95/00372
;      FILING DATE:  23-FEB-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9403423.8
;      FILING DATE:  23-FEB-1994
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  730 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-696-944-2

  Query Match             6.7%;  Score 83;  DB 2;  Length 730;
  Best Local Similarity   23.9%;  Pred. No. 0.87;
  Matches   50;  Conservative   31;  Mismatches   86;  Indels   42;  Gaps   10;

Qy          8 LGNSRFKCASL-------------REGVIGPVTVLPYLTETMDAFALQELPRGRVAYLA 53
              :||::    |:                ||:||||:    : | |      | : : :|
Db        538 VGNNKISLLSVSVGLANVGTHFETWNTGVLGPVTLTGLSSGTWD------LSKQKWSYKI 591

Qy         54 SVAAPAITTHVLEVLKIHFEQVQ--VAATVAACAGVRIAYAHPERFGVDRFLALLGSYGE 111
               :   ::: |  |      | ||  : |     |  :  :: |   | |    ||| |:
Db        592 GLKGESLSLHT-EAGSNSVEWVQGSLVAKKQPLAWYKTTFSAPA--GNDPLALDLGSMGK 648

Qy        112 GNVLVVG--VGTALTIDLLAAN--GCHLGGRISASPTL------MRQALHARAEQLPLSG 161
              | | | :|     :    |   |: |  : : |       ::  |     |  ||
Db        649 GEVWVNGQSIGRHWPGNKARGNCGNCNYAGTYTDTKCLANCGQPSQRWYHVPRSWL-RSG 707

Qy        162 GNYLEFAEDTEDALVSGCNGAAVALIERS 190
              ||||   |:       | :   :||:||:
Db        708 GNYLVVLEEW------GGDPNGIALVERT 730


RESULT 8
US-08-441-139-11
; Sequence 11, Application US/08441139
; Patent No. 5773245
;  GENERAL INFORMATION:
;    APPLICANT:  Wittrup, Dr. Karl D.
;    APPLICANT:  Robinson, Anne S.
;    TITLE OF INVENTION:  METHODS FOR INCREASING SECRETION OF
;    TITLE OF INVENTION:  RECOMBINANTLY EXPRESSED PROTEINS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,139
;      FILING DATE:  15-MAY-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/089,997
;      FILING DATE:  06-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8646
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  654 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-441-139-11

  Query Match             6.6%;  Score 81.5;  DB 1;  Length 654;
  Best Local Similarity   26.7%;  Pred. No. 1.1;
  Matches   36;  Conservative   17;  Mismatches   51;  Indels   31;  Gaps    7;

Qy         23 IGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAAT-- 80
              ||      :  | : |  | ::     |||         || :  :  :|   |  ||
Db        132 IGGGQTKTFAPEEISAMVLTKMKETAEAYLGKK-----VTHAVVTVPAYFNDAQRQATKD 186

Qy         81 VAACAGVRIAYAHPE------RFGVDRFLALLGSYGEGNVLVVGVG------TALTID-- 126
                 ||: :     |       :|:|:        || |:||  :|     : ||||
Db        187 AGTIAGLNVMRIINEPTAAAIAYGLDK------REGEKNILVFDLGGGTFDVSLLTIDNG 240

Qy        127 ---LLAANG-CHLGG 137
                ::| ||  ||||
Db        241 VFEVVATNGDTHLGG 255


RESULT 9
US-08-441-139-16
; Sequence 16, Application US/08441139
; Patent No. 5773245
;  GENERAL INFORMATION:
;    APPLICANT:  Wittrup, Dr. Karl D.
;    APPLICANT:  Robinson, Anne S.
;    TITLE OF INVENTION:  METHODS FOR INCREASING SECRETION OF
;    TITLE OF INVENTION:  RECOMBINANTLY EXPRESSED PROTEINS
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 Garden City Plaza
;      CITY:  Garden City
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/441,139
;      FILING DATE:  15-MAY-1995
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/089,997
;      FILING DATE:  06-JUL-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DiGiglio, Frank S.
;      REGISTRATION NUMBER:  31,346
;      REFERENCE/DOCKET NUMBER:  8646
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516-742-4343
;      TELEFAX:  516-742-4366
;      TELEX:  230 901 SANS UR
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  666 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-441-139-16

  Query Match             6.5%;  Score 80.5;  DB 1;  Length 666;
  Best Local Similarity   26.8%;  Pred. No. 1.5;
  Matches   34;  Conservative   17;  Mismatches   45;  Indels   31;  Gaps    7;

Qy         31 YLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAAT--VAACAGVR 88
              :  | : |  | ::     |||         || :  :  :|   |  ||      ||:
Db        152 FAPEEISAMVLTKMKETAEAYLGKK-----VTHAVVTVPAYFNDAQRQATKDAGTIAGLN 206

Qy         89 IAYAHPE------RFGVDRFLALLGSYGEGNVLVVGVG------TALTID-----LLAAN 131
              :     |       :|:|:        || |:||  :|      : ||||     ::| |
Db        207 VMRIINEPTAAAIAYGLDK------REGEKNILVFDLGGGTFDVSLLTIDNGVFEVVATN 260

Qy        132 G-CHLGG 137
              |  ||||
Db        261 GDTHLGG 267


RESULT 10
US-09-044-718-82
; Sequence 82, Application US/09044718
; Patent No. 6391605
; GENERAL INFORMATION:
;  APPLICANT:  KOSTREWA, Dirk
;  APPLICANT:  PASAMONTES, Luis
;  APPLICANT:  TOMSCHY, Andrea
;  APPLICANT:  van LOON, Adolphus
;  APPLICANT:  VOGEL, Kurt
;  APPLICANT:  WYSS, Markus
;  TITLE OF INVENTION:  MODIFIED PHYTASES
;  FILE REFERENCE:  Modified Phytases
;  CURRENT APPLICATION NUMBER: US/09/044,718
;  CURRENT FILING DATE:  1998-03-19
;  PRIOR APPLICATION NUMBER:  EP 97810175.6
;  PRIOR FILING DATE:  1997-03-25
;  NUMBER OF SEQ ID NOS:  82
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 82
;   LENGTH: 469
;   TYPE: PRT
;   ORGANISM: Aspergillus fumigatus
US-09-044-718-82

  Query Match             6.5%;  Score 80;  DB 4;  Length 469;
  Best Local Similarity   24.9%;  Pred. No. 0.98;
  Matches   48;  Conservative   17;  Mismatches   72;  Indels   56;  Gaps    9;

Qy         39 FALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQ----------------VAATV 81
              |   ||      |   :: ||||   :    : |   ||                || |
Db        220 FEASELGDEVEANFTALFAPAIRARI----EKHLPGVQLTDDDVVSLMDMCSFDTVARTA 275

Qy         82 AA------CAGVRIAYAHPERFGVDRFLALLG---SYGEGNVLVVGVGTALTIDLLAANG 132
               |      ||     : | |    | :|  ||    || || |     |   | :|:|
Db        276 DASELSPFCA----IFTHNEWKKYD-YLQSLGKYYGYGAGNPLGPAQGIGFTNELIA--- 327

Qy        133 CHLGGRISASP----TLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIE 188
                    |:: ||     |      | :      ||:   |::|: |          ||
Db        328 -----RLTNSPVQDHTSTNSTLDSDPATFPLNATIYVDFSHD---------NGMIPIFFA 373

Qy        189 RSLYEAHQRLDQS 201
                ||   : | |:
Db        374 MGLYNGTEPLSQT 386


RESULT 11
US-09-199-637A-273
; Sequence 273, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
;  APPLICANT:  Ausubel, Frederick
;  APPLICANT:  Goodman, Howard M.
;  APPLICANT:  Rahme, Laurence G.
;  APPLICANT:  Mahajan-Miklos, Shalina
;  APPLICANT:  Tan, Man-Wah
;  APPLICANT:  Cao, Hui
;  APPLICANT:  Drenkard, Eliana
;  APPLICANT:  Tsongalis, John
;  TITLE OF INVENTION:  VIRULENCE-ASSOCIATED NUCLEIC ACID
;  TITLE OF INVENTION:  SEQUENCES AND USES THEREOF
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE:  1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 273
;   LENGTH: 989
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-273

  Query Match             6.5%;  Score 80;  DB 4;  Length 989;
  Best Local Similarity   29.1%;  Pred. No. 3.2;
  Matches   48;  Conservative   16;  Mismatches   59;  Indels   42;  Gaps    8;

Qy        103 LALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPT----LMRQALHARAEQLP 158
              || ||: |     | ||        ||      |   | :|:      | | | || |||
Db        808 LAFLGALGH----VDGVLVQRLARLLDVRVVH---RFAATQVGHGILQRLARHALAEQQL 860

Qy        159 LSGGNYLEFAEDTE---DALVSGCNGAAVALIERS---LYEAH---------QRLDQSVR 203
                  |  :   :  :   | ||:   | ||:|:|::   | : |          ||:: |
Db        861 AEPGVLVHRGQQYQLAGDELVALLLGQAVSLVEQACEILGQVHVAGRALDLRQRVEFFVE 920

Qy        204 LLLHGGGV------------ASLLPWLGDVVHRPTLVLDGLAIWA 236
                  || :             | ||   |  |||    |||  : |
Db        921 AAAQGGDIEADLHQQGLDRTALLLEQGGKQVHR----LDGRMVMA 961


RESULT 12
US-09-413-814-105
; Sequence 105, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
;  APPLICANT: Dougherty, Brian A
;  APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
```

```
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 105
;   LENGTH: 829
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-105

  Query Match             6.4%;  Score 79;  DB 4;  Length 829;
  Best Local Similarity   24.3%;  Pred. No. 3.1;
  Matches   72;  Conservative   30;  Mismatches   88;  Indels  106;  Gaps   15;

Qy         16 ASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAY----------------LAS----- 54
              |:| ||  |  ||           | : || |:||:                 |||
Db        314 AALAEGRAGAGTV-----------AGEALPPGKVAFVFPGQGSQWPSMARALLASSPAFR 362

Qy         55 --------VAAPAITTHVLEVL---KIH----FEQVQV------AATVAACAGVRIAYAH 93
                    || :   :| ||   : |      |:| |      |  :|  |  | |
Db        363 AEIEACERALAPHVDWSLLAVLGGDEAHAAPMLERVDVVQPVLFAVMIALAATWRAAGVT 422

Qy         94 PERFGVDRFLALLGSYGEGNVLVVGVGTALTID-------LLAANGCHLGGRISASPTLM 146
              |:       |::| : :| :    |  ||:::       | :     | || : |   :
Db        423 PD--------AVVG-HSQGEIAAAYVAGALSLEDAARVVALRSRAITKLAGRGAMSAVEL 473

Qy        147 RQA-LHARAEQLPLSGGNYLEFA--EDTEDALVSGCNGAAVALIE-----------RSL 191
                | | ||   |   |  | |       |||||  ||   |:             |
Db        474 TTAELEARLAPL----GERLAIAAINSPHAALVSGDPGAIDELVAELSGAQLFARKVRVE 529

Qy        192 YEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLV----------LDGLAIWAA 237
              | :|    :::          ::|  |  :  ||  |          ||| |: ||
Db        530 YASHSAHVEAIE--------RTMLEALDGIAPRPATVPLYSAVTGERLDGEALGAA 577


RESULT 13
US-09-413-814-48
; Sequence 48, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
;  APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
;  APPLICANT: Bristol-Myers Squibb, Co.
;  APPLICANT: Beyer, Stefan
;  APPLICANT: Bloecker, Helmut
;  APPLICANT: Brandt, Petra
;  APPLICANT: Cino, Paul M
   APPLICANT: Dougherty, Brian A
   APPLICANT: Goldberg, Steven L
;  APPLICANT: Hofle, Gerhard
;  APPLICANT: Mueller, Joachim
;  APPLICANT: Reichenbach, Hans
;  TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
;  TITLE OF INVENTION: heteropolyketide compounds
;  FILE REFERENCE: PCT/US 99/23535
;  CURRENT APPLICATION NUMBER: US/09/413,814
;  CURRENT FILING DATE: 1999-10-07
;  EARLIER APPLICATION NUMBER: DE 198 46 493.2
;  EARLIER FILING DATE: 1998-10-09
;  NUMBER OF SEQ ID NOS: 107
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 48
;   LENGTH: 2475
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-413-814-48

  Query Match             6.4%;  Score 79;  DB 4;  Length 2475;
  Best Local Similarity   23.7%;  Pred. No. 18;
  Matches   61;  Conservative   20;  Mismatches   84;  Indels   92;  Gaps   13;

Qy         25 PVTVLPYLTETMDAFALQELPRGRVAYLA---SVAAPAITTHVLEVLKIHFEQVQVAATV 81
              | |  |  ||  |      |  |     |    |  |   |:   |  | |:
Db       1391 PTTAAPLSTEPND------LDAGDADGLVLPISARTPTALAHIATNLANHLER------- 1437

Qy         82 AACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISA 141
                        ||        :||                 |||: |       |    |
Db       1438 ----------HPT-------IAL------------ADVALTLQLGRRQWPHRHSLICR 1466

Qy        142 SPT----LMRQALHARAE----QLPLS-----------GGNYLEFAEDTEDALVSGCNG 181
              : |    |:|  :|: ||    | |:|           |  |   | |    ||  |
Db       1467 NRTEAIKLLRAVVHS-AEVPPAQAPVSDAPRCVFLFPGQGAQYPSMARD----LVRNCPD 1521

Qy        182 AAVALIERSLYEAHQRLDQSVRLLLHGGGVA----------------SLLPWLGDVVH 223
               |: | :  | :  : | :  | :| | | |                   :|  ||||
Db       1522 FALHL-DPCLDQLAELLPEDPRCILFGDGPADRLDQTAYTQPLLFSVSYALARWLGDFGI 1580

Qy        224 RPTLVLDGLAIWAAVAA 240
              ||  :: | ::   |||
Db       1581 RPDAMI-GHSLGEYVAA 1596


RESULT 14
US-08-801-344-8
; Sequence 8, Application US/08801344
; Patent No. 6087140
;  GENERAL INFORMATION:
;    APPLICANT:  Cameron, Douglas C.
;    APPLICANT:  Shaw, Anita J.
;    APPLICANT:  Altaras, Nedim E.
;    TITLE OF INVENTION:  MICROBIAL PRODUCTION OF 1,2-PROPANEDIOL
;    TITLE OF INVENTION:  FROM SUGAR
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DeWitt Ross & Stevens S.C.
;      STREET:  8000 Excelsior Drive, Suite 401
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53717-1914
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/801,344
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REFERENCE/DOCKET NUMBER:  09820.037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-831-2100
;      TELEFAX:  608-831-2106
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  510 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Pyridine nucleotide transhydrogenase, subunit
;      ORGANISM:  A
US-08-801-344-8

  Query Match             6.4%;  Score 78.5;  DB 3;  Length 510;
  Best Local Similarity   22.0%;  Pred. No. 1.7;
  Matches   48;  Conservative   27;  Mismatches   94;  Indels   49;  Gaps    8;
```

```
Qy          27 TVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAG 86
               |:: ::    :   :|:|    |  :|  : | |:          : :   :::|  ||
Db          88 TLVSFIWPAQNPELMQKLAERNVTVMAMDSVPRIS---------RAQSLDALSSMANIAG 138

Qy          87 VRIAYAHPERFGVDRF----LALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISA- 141
                |        ||  ||    :   |      |:|:| | |    : |||   ||  : |
Db         139 YRAIVEAAHEFG--RFFTGQITAAGKVPPAKVMVIGAGVAGLAAIGAANS--LGAIVRAF 194

Qy         142 --SPTLMRQALHARAEQLPL-------SGGNYLEFAED-------------------TE 172
                  | :  |     || | |        ||  | :   |
Db         195 DTRPEVKEQVQSMGAEFLELDFKEEAGSGDGYAKVMSDAFIKAEMELFAAQAKEVDIIVT 254

Qy         173 DALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGG 210
                ||: |    |  || | : :: :     | |    ||
Db         255 TALIPG--KPAPKLITREMVDSMKAGSVIVDLAAQNGG 290
```

```
RESULT 15
US-09-498-599-8
; Sequence 8, Application US/09498599
; Patent No. 6303352
;  GENERAL INFORMATION:
;    APPLICANT:  Cameron, Douglas C.
;    APPLICANT:  Shaw, Anita J.
;    APPLICANT:  Altaras, Nedim E.
;    TITLE OF INVENTION:  MICROBIAL PRODUCTION OF
;    TITLE OF INVENTION:  1,2-PROPANEDIOL FROM SUGAR
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DeWitt Ross & Stevens S.C.
;      STREET:  8000 Excelsior Drive, Suite 401
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  U.S.A.
;      ZIP:  53717-1914
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/498,599
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sara, Charles S.
;      REFERENCE/DOCKET NUMBER:  09820.037
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-831-2100
;      TELEFAX:  608-831-2106
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  510 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  N-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Pyridine nucleotide transhydrogenase, subunit
;      ORGANISM:  A
US-09-498-599-8

  Query Match             6.4%;  Score 78.5;  DB 4;  Length 510;
  Best Local Similarity   22.0%;  Pred. No. 1.7;
  Matches   48;  Conservative   27;  Mismatches   94;  Indels   49;  Gaps    8;

Qy          27 TVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAG 86
               |:: ::    :   :|:|    |  :|  : | |:          : :   :::|  ||
Db          88 TLVSFIWPAQNPELMQKLAERNVTVMAMDSVPRIS---------RAQSLDALSSMANIAG 138

Qy          87 VRIAYAHPERFGVDRF----LALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISA- 141
                |        ||  ||    :   |      |:|:| | |    : |||   ||  : |
Db         139 YRAIVEAAHEFG--RFFTGQITAAGKVPPAKVMVIGAGVAGLAAIGAANS--LGAIVRAF 194

Qy         142 --SPTLMRQALHARAEQLPL-------SGGNYLEFAED-------------------TE 172
                  | :  |     || | |        ||  | :   |
Db         195 DTRPEVKEQVQSMGAEFLELDFKEEAGSGDGYAKVMSDAFIKAEMELFAAQAKEVDIIVT 254

Qy         173 DALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGG 210
                ||: |    |  || | : :: :     | |    ||
Db         255 TALIPG--KPAPKLITREMVDSMKAGSVIVDLAAQNGG 290

Search completed: June 24, 2003, 22:10:52
Job time : 10.8186 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:24:12 ; Search time 13.449 Seconds
(without alignments)
1947.059 Million cell updates/sec

Title: US-09-813-453A-65
Perfect score: 1230
Sequence: 1 MNDWLFDLGNSRFKCASLRE.....HRPTLVLDGLAIWAAVAANV 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1230	100.0	242	9	US-09-813-453A-65	Sequence 65, Appl
2	200	16.3	592	9	US-09-813-453A-43	Sequence 43, Appl
3	199.5	16.2	460	9	US-09-813-453A-39	Sequence 39, Appl
4	198.5	16.1	592	9	US-09-813-453A-22	Sequence 22, Appl
5	194.5	15.8	267	9	US-09-813-453A-15	Sequence 15, Appl
6	193	15.7	248	9	US-09-813-453A-20	Sequence 20, Appl
7	181.5	14.8	244	9	US-09-813-453A-41	Sequence 41, Appl
8	168.5	13.7	249	9	US-09-813-453A-61	Sequence 61, Appl
9	165.5	13.5	249	9	US-09-813-453A-70	Sequence 70, Appl
10	144	11.7	257	9	US-09-813-453A-53	Sequence 53, Appl
11	143.5	11.7	219	9	US-09-813-453A-57	Sequence 57, Appl
12	143.5	11.7	241	9	US-09-813-453A-63	Sequence 63, Appl
13	140.5	11.4	260	9	US-09-813-453A-51	Sequence 51, Appl
14	139	11.3	265	9	US-09-813-453A-4	Sequence 4, Appli
15	134.5	10.9	258	9	US-09-813-453A-2	Sequence 2, Appli
16	128.5	10.4	262	9	US-09-813-453A-8	Sequence 8, Appli
17	128	10.4	255	9	US-09-813-453A-7	Sequence 7, Appli
18	125.5	10.2	258	9	US-09-813-453A-6	Sequence 6, Appli
19	120	9.8	256	9	US-09-813-453A-55	Sequence 55, Appl
20	117	9.5	262	9	US-09-813-453A-45	Sequence 45, Appl
21	116	9.4	233	9	US-09-813-453A-17	Sequence 17, Appl
22	115.5	9.4	223	9	US-09-895-913A-74	Sequence 74, Appl
23	115.5	9.4	223	9	US-09-813-453A-14	Sequence 14, Appl
24	115.5	9.4	223	9	US-09-813-453A-67	Sequence 67, Appl
25	115	9.3	258	9	US-09-813-453A-49	Sequence 49, Appl
26	112	9.1	272	9	US-09-712-363-276	Sequence 276, App
27	112	9.1	272	9	US-09-813-453A-5	Sequence 5, Appli
28	108.5	8.8	250	9	US-09-813-453A-3	Sequence 3, Appli
29	107.5	8.7	254	9	US-09-813-453A-47	Sequence 47, Appl
30	105	8.5	498	10	US-09-815-242-11777	Sequence 11777, A
31	104	8.5	212	9	US-09-813-453A-59	Sequence 59, Appl
32	103	8.4	246	9	US-09-813-453A-9	Sequence 9, Appli
33	99	8.0	257	9	US-09-813-453A-13	Sequence 13, Appl
34	95	7.7	229	9	US-09-813-453A-12	Sequence 12, Appl
35	93.5	7.6	262	9	US-09-813-453A-11	Sequence 11, Appl
36	93.5	7.6	323	10	US-09-815-242-13734	Sequence 13734, A
37	93.5	7.6	4999	9	US-09-976-059-15	Sequence 15, Appl
38	89.5	7.3	209	9	US-09-813-453A-21	Sequence 21, Appl
39	88	7.2	273	9	US-09-813-453A-10	Sequence 10, Appl
40	81.5	6.6	653	10	US-09-759-010-2	Sequence 2, Appli
41	81.5	6.6	654	9	US-09-919-039-260	Sequence 260, App
42	81.5	6.6	654	10	US-09-919-172-54	Sequence 54, Appl
43	81.5	6.6	4999	9	US-09-976-059-14	Sequence 14, Appl
44	80.5	6.5	323	10	US-09-815-242-10321	Sequence 10321, A
45	80.5	6.5	342	9	US-09-712-363-210	Sequence 210, App

ALIGNMENTS

RESULT 1
US-09-813-453A-65
; Sequence 65, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 242
; TYPE: PRT
; ORGANISM: Xylessa fastidiosa
US-09-813-453A-65

Query Match 100.0%; Score 1230; DB 9; Length 242;
Best Local Similarity 100.0%; Pred. No. 7.3e-107;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60

Qy         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120

Qy        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180

Qy        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240

Qy        241 NV 242
              ||
Db        241 NV 242


RESULT 2
US-09-813-453A-43
; Sequence 43, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
   PRIOR FILING DATE: 2000-08-24
   PRIOR APPLICATION NUMBER: 09/667,569
   PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-43

  Query Match            16.3%;  Score 200;  DB 9;  Length 592;
  Best Local Similarity  30.6%;  Pred. No. 2e-10;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | |  :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571

Qy        231 GLAIW 235
              | |:
Db        572 NLVIY 576


RESULT 3
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39

  Query Match            16.2%;  Score 199.5;  DB 9;  Length 460;
  Best Local Similarity  30.3%;  Pred. No. 1.6e-10;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | : |||| | | :  |    |   ||   :       |   | |  :  :|   :
Db        210 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 269

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        270 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 326

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        327 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 386

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        387 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 442

Qy        229 LDGL 232
              : ||
Db        443 IHGL 446


RESULT 4
US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-22

  Query Match            16.1%;  Score 198.5;  DB 9;  Length 592;
  Best Local Similarity  30.7%;  Pred. No. 2.7e-10;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | |  :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
```

```
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578
```

RESULT 5
US-09-813-453A-15
; Sequence 15, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Bordetella pertussis
US-09-813-453A-15

Query Match 15.8%; Score 194.5; DB 9; Length 267;
Best Local Similarity 27.9%; Pred. No. 2.3e-10;
Matches 80; Conservative 25; Mismatches 105; Indels 77; Gaps 10;

```
Qy          5 LFDLGNSRFKCA---------SLREGVIGPVTVLPYLTETMDAFA----LQELPRG----- 47
              | | |||| |           : ||         |   : :|  |    |  |||
Db          4 LIDSGNSRLKVGWFDPDAPQAAREPA-------PVAFDNLDLDALGRWLATLPRRPQRAL 56

Qy         48 --RVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAA-------------CAGVRIAYA 92
                || ||   |                    :|||: |               |:|  |
Db         57 GVNVAGLARGEA------------------IAATLRAGGCDIRWLRAQPLAMGLRNGYR 97

Qy         93 HPERFGVDRFLALLGSYG-----EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMR 147
              :|:: | ||:  ::|           :||   ||| |:| :  :    || |   | :||
Db         98 NPDQLGADRWACMVGVLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMR 157

Qy        148 QALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLH 207
               ||      |||: |   ::  ||  |: ||   |    | |      ||  |:  : :
Db        158 GALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQAGAIVRQWLAGRQRYGQAPEIYVA 217

Qy        208 GGG-------VASLLPWLG---DVVHRPTL----VLDGLAIWAAVAA 240
              |||          ||   |       :||     ||||||  ||  |
Db        218 GGGWPEVRQEAERLLAVTGAAFGATPQPTYLDSPVLDGLAALAAQGA 264
```

RESULT 6
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

Query Match 15.7%; Score 193; DB 9; Length 248;
Best Local Similarity 31.0%; Pred. No. 2.9e-10;
Matches 76; Conservative 34; Mismatches 111; Indels 24; Gaps 8;

```
Qy          7 DLGNSRFKCASLREGVIGPVTVLPYLTETMDAF-----ALQELPRGRVAYLASVAAPAIT 61
              | |||  |   : ||     :|   | |: ||      : | ||  |   | || :   |
Db          6 DCGNSLIKWRVI-EGAAR--SVAGGLAESDDALVEQLTSQQALP-VRACRLVSVRSEQET 61

Qy         62 THVLEVLKIHFEQVQVAATVAA----CAGVRIAYAHPERFGVDRFLALLGSY--GEGNVL 115
              : ::  |: |    |:| ||:     ||||  |   :| |:||:|||: ::   :   |
Db         62 SQLVARLEQLF---PVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACL 118

Qy        116 VVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN------YLEFAE 169
              |: :|||:| ||:||:| |||| |    ||||  |    ::           |:  :
Db        119 VIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQ 178

Qy        170 DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVL 229
              | :|:  ||       :      | : |     : | ||    :   |      | ||
Db        179 ATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVF 238

Qy        230 DGLAI 234
                |||:
Db        239 VGLAL 243
```

RESULT 7
US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

Query Match 14.8%; Score 181.5; DB 9; Length 244;
Best Local Similarity 28.0%; Pred. No. 3.3e-09;
Matches 72; Conservative 39; Mismatches 91; Indels 55; Gaps 12;

```
Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAY-LASVAAP----- 58
              : | |||  | | :|   |  :: ||                |:    |: : ||
Db          6 IVDQGNSACKVAFVRNNSIESISFLP---------------GKAGQALSHLVAPHRFDK 49

Qy         59 AITTHVLEVLKIHFEQVQVAATVAACAG-----------VRIAYAHPERFGVDRFLALL 106
              || :    :   : | :||             :|: |      | ||  |::
Db         50 AIYS------SVGLPDEEAEAIVRSCAAASLMMGTETPVPLRLQYDR-RTLGADRLAAVV 102

Qy        107 GS---YGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPL---S 160
              |:   |    :||:  |||:| : ::| | :||| ||    |  :|||    :||| |
Db        103 GAHSLYPNTELLVIDAGTAITYERVSAEGIYLGGNISPGLHLRFKALHLFTGRLPLIDPS 162
```

```
Qy        161 G--GNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWL 218
              |      |:   ||:|: :|        |:| : : | :  :|  :|   || |: |  :
Db        163 GISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSAVILT--GGDANYLARI 220

Qy        219 ---GDVVHRPTLVLDGL 232
                 | ::| | ||| ||
Db        221 IRSGILIH-PDLVLLGL 236


RESULT 8
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
   PRIOR APPLICATION NUMBER: US 60/227,860
   PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             13.7%;  Score 168.5;  DB 9;  Length 249;
  Best Local Similarity   29.7%;  Pred. No. 5.6e-08;
  Matches   74;  Conservative   34;  Mismatches  110;  Indels   31;  Gaps   10;

Qy          7 DLGNS--RFKCASLREGVI---GPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAIT 61
              | |||  :::   : : ||   | |     |  : | |   |   |:    || :  |
Db          6 DCGNSFIKWRVIHVADAVIEGGGIVDSDQALVAEVAALASVRLTGCRI---VSVRSEEET 62

Qy         62 THVLEVLKIHFE-QVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSY--GEGNVLVVG 118
                :  ::   |  | :||  |   |||| |   :| |:||:|| ||::   :|  ||:
Db         63 DALCALIAQAFAVQAKVAHPVREMAGVRNGYDDYQRLGMDRWLAALGAFHLAKGACLVID 122

Qy        119 VGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGG------NYLEFAEDTE 172
              :|||   | ::|:| |||| |     |||  |     ::           : |     |
Db        123 LGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRIRYDDASAERALSSLSPGRSTV 182

Qy        173 DALVSGC----NGAAVALIERSLYEAHQRLDQSVRLLLHGGG---VASLLPWLGDVVHRP 225
              :|:  ||     | |   :|    :|     :   : | ||    | : ||    |   |
Db        183 EAVERGCVLMLQGFAYTQLE----QARVLWGEEFTVFLTGGDAPLVRAALPQARVV---P 235

Qy        226 TLVLDGLAI 234
                ||  |||:
Db        236 DLVFVGLAM 244


RESULT 9
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             13.5%;  Score 165.5;  DB 9;  Length 249;
  Best Local Similarity   29.8%;  Pred. No. 1.1e-07;
  Matches   75;  Conservative   29;  Mismatches  111;  Indels   37;  Gaps   10;

Qy          7 DLGNSRFK-------CASL-REGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAP 58
              | |||  |       |::|   ||:   | |      :   |  :         | || :
Db          6 DCGNSFIKWRIITKSCSTLVSGGVVDSDTALLECLGNLSGAAFSD------CRLVSVRSA 59

Qy         59 AITTHVLEVLKIHFEQVQV-AATVAACAGVRIAYAHPERFGVDRFLALLGSYG--EGNVL 115
                |  ::  |   |    | ||     |||   |    | |:||:|| :|:|   :|  |
Db         60 EETAKLVCALADTFSISPVCAAPAPELAGVINGYDDFARLGLDRWLAFVGAYHLVKGACL 119

Qy        116 VVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN------YLEFAE 169
              |: :|||:| | : |:| |||| |     |||  |     ::           |
Db        120 VIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEAEKALVRLVPGR 179

Qy        170 DTEDALVSGCN----GAAVALIERSLYEAHQRLDQSVRLLLHGGG---VASLLPWLGDVV 222
                | :|:  ||:    | |:  ||     | :      : : ||    || :||    |
Db        180 ATAEAVERGCSLMLRGFAMTQIE----IAREYWGDDFAIFVTGGDAVLVADVLPGARIV- 234

Qy        223 HRPTLVLDGLAI 234
                | ||  |||:
Db        235 --PDLVFVGLAL 244


RESULT 10
US-09-813-453A-53
; Sequence 53, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Chlorobium tepidum
US-09-813-453A-53

  Query Match             11.7%;  Score 144;  DB 9;  Length 257;
  Best Local Similarity   29.2%;  Pred. No. 1.1e-05;
  Matches   57;  Conservative   26;  Mismatches   82;  Indels   30;  Gaps    8;

Qy         56 AAPAITTHVLEVLKIHFEQVQVAATVAACA---GVRIAYAHPERFGVDRFLALLG----S 108
              :| |: : :|| |         |       |       |: || |  || || |||
Db         64 SATAVGSALLESL------FSVPVLTICCKLRFPFRLDYATPHTFGADR-LALCAWSRHL 116

Qy        109 YGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLE-- 166
              : |  |: | :|||:| |:|   | : || |     :|  |||:|  |||    :  |
Db        117 FSEKPVIAVDIGTAITFDVLDTVGNYRGGLIMPGIDMMAGALHSRTAQLPQVRIDRPESL 176

Qy        167 FAEDTEDALVSGC-------NGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLG 219
                  | : ||          |  |  |  |        : :| ::: ||    ::| :|
Db        177 LGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFG--ESTVEVIVTGGNSRIIVPEIG 234
```

```
Qy        220 DVVHRPTLVLDGLAI 234
                  |  |:| ||:
Db        235 -----PVSVIDELAV 244
```

RESULT 11
US-09-813-453A-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: US 60/227,860
PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Dehalococcoides ethenogenes
US-09-813-453A-57

Query Match 11.7%; Score 143.5; DB 9; Length 219;
Best Local Similarity 25.6%; Pred. No. 1e-05;
Matches 51; Conservative 29; Mismatches 86; Indels 33; Gaps 6;

```
Qy         16 ASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQV 75
                |||  |::   : |          :|| |     : ||  | :||   || | :|:
Db         43 ASLLLGLLQHAGIHP----------EELNR---VIMCSV-VPPLTTTFEEVFKSYFKAA 87

Qy         76 QVAATVAACAGVRIAYAHPERFGVDRFL--ALLGSYGEGNVLVVGVGTALTIDLLAANGC 133
               :      :||::   :|   | || :  |       |  ::| :||| | | |:  |
Db         88 PLVVGAGIKSGVKVRMDNPREVGADRIVNAAAARVLYPGACIIVDMGTATTFDTLSEGGA 147

Qy        134 HLGGRISASPTLMRQALHARAEQLP---------LSGGNYLEFAEDTEDALVSGCNGAAV 184
               ::|| |:       ||:  :  :||         : | |       |  |: ||     :
Db        148 YIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAKVIGSN-------TVSAMQSGIYFGYI 200

Qy        185 ALIERSLYEAHQRLDQSVR 203
                |:|  :      | |  |
Db        201 GLVEELVRRIQTELGQKTR 219
```

RESULT 12
US-09-813-453A-63
; Sequence 63, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 63
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Thiobacillus ferrooxidans
US-09-813-453A-63

Query Match 11.7%; Score 143.5; DB 9; Length 241;
Best Local Similarity 27.2%; Pred. No. 1.1e-05;
Matches 70; Conservative 31; Mismatches 103; Indels 53; Gaps 11;

```
Qy          8 LGNSRFKCASLREGV----IGPVTVLPYLTETMDAFALQEL--PRGRVAYLASVAAPAI- 60
              :||:|   |  . :||    :   | ||  || :    |  |  |      |  |   |:
Db          6 VGNTRTLLAHTHDGVHFDSVSVATSLP-PTEILQQPGLTWLSAPNREPVALGGVVPAALA 64

Qy         61 -------TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS---YG 110
                    |  | |     | :   \         |   |  ||  | ||   || :   |
Db         65 AWREALATAEVREPDPGFFRR----------AVPHDYHPPESLGFDRRCCLLAAAMDYP 113

Qy        111 EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH----------ARAEQLP 158
                : :|: :|||:|||| | |   ||||     :  : ||           | || |
Db        114 GQDSIVIDMGTAITIDLL-AGGHFRGGRILPGIAMSLRGLHEGTALLPEVVLNAPAEML- 171

Qy        159 LSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWL 218
                 ||      || :|: :|        :  :: :  |   |: |:|: ||      | :
Db        172 ---GN------DTSNAIQAGVIHLFADALRGAITDFRQYSPQA-RILITGGDAERWQPGI 221

Qy        219 GDVVHRPTLVLDGLAIW 235
                :::| |:| |  :|
Db        222 AGSLYQPHLLLRGFYLW 238
```

RESULT 13
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-09-813-453A-51

Query Match 11.4%; Score 140.5; DB 9; Length 260;
Best Local Similarity 27.2%; Pred. No. 2.4e-05;
Matches 49; Conservative 30; Mismatches 92; Indels 9; Gaps 4;

```
Qy         50 AYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS- 108
              | : |   |    ::  : : :|    :      |  |: :    |   | || :  :|:
Db         59 AVIISSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAA 118

Qy        109 -YGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP-----LSGG 162
                 | ::|:  ||| | |::||:|   || |:    |  ||||  | :||       |
Db        119 MVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAG 178

Qy        163 NYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV 222
              | :    ||  |: ||     ::|||  :        : : ::   ||||||     | :
Db        179 NRI-VGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIA-TGGVASLFEGATDSI 236
```

RESULT 14
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:

```
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

  Query Match             11.3%;  Score 139;  DB 9;  Length 265;
  Best Local Similarity   26.5%;  Pred. No. 3.4e-05;
  Matches   62;  Conservative   33;  Mismatches  107;  Indels   32;  Gaps    9;

Qy         16 ASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQV 75
              | | :|::|           |      ||  |          |::   : || : ::  |
Db         40 AVLLQGLMG----------MHPLLGDELGDGIDGIAICATVPSVLHELREVTRRYYGDV 88

Qy         76 -QVAATVAACAGVRIAYAHPERFGVDRFL---ALLGSYGEGNVLVVGVGTALTIDLLAAN 131
               |       || |   ||:  | || :   | :  || |  :||  ||| | | ::|
Db         89 PAVLVEPGVKTGVPILTDHPKEVGADRIINAVAAVELYG-GPAIVVDFGTATTFDAVSAR 147

Qy        132 GCHLGGRISASPTLMRQALHARAEQL-------PLS--GGNYLEFAEDTEDALVSGCNGA 182
              | ::|| |:    :  :||  :  ||       | |  | | :   |  :  :| |  |
Db        148 GEYIGGVIAPGIEISVEALGVKGAQLRKIEVARPRSVIGKNTV---EAMQSGIVYGFAGQ 204

Qy        183 AVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV--HRPTLVLDGLAI 234
                 :: |   |     |     ::  ||:| ::     |:  | | | | || :
Db        205 VDGVVNRMARELADDPDDVT--VIATGGLAPMVLGESSVIDEHEPWLTLMGLRL 256


RESULT 15
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
   FILE REFERENCE: OGZ-001
   CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 258
;   TYPE: PRT
;   ORGANISM: Bacillus subtilis
US-09-813-453A-2

  Query Match             10.9%;  Score 134.5;  DB 9;  Length 258;
  Best Local Similarity   23.8%;  Pred. No. 8.5e-05;
  Matches   58;  Conservative   40;  Mismatches   91;  Indels   55;  Gaps    9;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHV 64
              |||     |:    :|:|    | | :      ||: :          :  | : |
Db         46 LFDHSGLMFEQI---DGIIISSVVPPIM------FALERMCTKYFHIEPQIVGPGMKT-- 94

Qy         65 LEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRF---LALLGSYGEGNVLVVGVGT 121
                                  |: | | :|:  | ||    :| :  ||   ::|| ||
Db         95 --------------------GLNIKYDNPKEVGADRIVNAVAAIHLYG-NPLIVVDFGT 132

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP---------LSGGNYLEFAEDTE 172
              | |   :  |  ::|| |:   |:  :||::|| :||         : | | :      :
Db        133 ATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNTV---SAMQ 189

Qy        173 DALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVH--RPTLVLD 230
              ::|  |    :::|  ::| | |      ::  ||:| |:    | :    | | |
Db        190 SGILFGYVGQVEGIVKRMKWQAKQDLK-----VIATGGLAPLIANESDCIDIVDPFLTLK 244

Qy        231 GLAI 234
                || :
Db        245 GLEL 248


Search completed: June 24, 2003, 22:29:09
Job time : 13.449 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:59:48 ; Search time 21.2874 Seconds
(without alignments)
2342.388 Million cell updates/sec

Title: US-09-813-453A-65
Perfect score: 1230
Sequence: 1 MNDWLFDLGNSRFKCASLRE.....HRPTLVLDGLAIWAAVAANV 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1230	100.0	242	16	Q9PCI4	Q9pci4 xylella fas
2	214.5	17.4	295	16	Q8Y2M4	Q8y2m4 ralstonia s
3	200	16.3	592	16	Q9JXF1	Q9jxf1 neisseria m
4	198.5	16.1	592	16	Q9JWI7	Q9jwi7 neisseria m
5	193	15.7	248	16	Q9HWC1	Q9hwc1 pseudomonas
6	140.5	11.4	261	16	Q9A6Z1	Q9a6z1 caulobacter
7	139	11.3	265	16	Q9X8N6	Q9x8n6 streptomyce
8	130	10.6	273	16	Q97EB4	Q97eb4 clostridium
9	128.5	10.4	262	16	Q9RX54	Q9rx54 deinococcus
10	122.5	10.0	259	16	Q8XHL5	Q8xhl5 clostridium
11	119	9.7	255	16	Q8R7M2	Q8r7m2 thermoanaer
12	117	9.5	258	2	Q9F985	Q9f985 bacillus st
13	115.5	9.4	223	16	O25533	O25533 helicobacte
14	113.5	9.2	274	16	Q9CD56	Q9cd56 mycobacteri
15	112	9.1	272	16	O06282	O06282 mycobacteri
16	107.5	8.7	223	16	Q9ZKY6	Q9zky6 helicobacte
17	107.5	8.7	254	16	Q9KGH5	Q9kgh5 bacillus ha
18	106	8.6	259	16	Q8YAC5	Q8yac5 listeria mo
19	104.5	8.5	276	16	Q8YQD7	Q8yqd7 anabaena sp
20	104	8.5	212	2	O32514	O32514 desulfovibr
21	103	8.4	246	16	Q9WZY5	Q9wzy5 thermotoga
22	101	8.2	259	16	Q92F54	Q92f54 listeria in
23	99.5	8.1	241	16	Q93J98	Q93j98 streptomyce
24	99	8.0	257	16	P74045	P74045 synechocyst
25	97.5	7.9	438	16	Q8Y1J7	Q8y1j7 ralstonia s
26	95.5	7.8	464	16	Q9RW77	Q9rw77 deinococcus
27	95	7.7	229	16	O67753	O67753 aquifex aeo
28	94.5	7.7	600	17	Q8TWQ6	Q8twq6 methanopyru
29	94	7.6	376	16	Q8XRA9	Q8xra9 ralstonia s
30	93.5	7.6	262	16	O51477	O51477 borrelia bu
31	93.5	7.6	323	16	Q8XFR7	Q8xfr7 salmonella
32	93.5	7.6	379	16	Q8XVU7	Q8xvu7 ralstonia s
33	93	7.6	733	16	Q8YB89	Q8yb89 brucella me
34	93	7.6	1245	16	Q8XYE3	Q8xye3 ralstonia s
35	93	7.6	4077	2	Q939Z0	Q939z0 amycolatops
36	92.5	7.5	631	5	Q25325	Q25325 leishmania
37	91.5	7.4	433	16	Q8Y1M4	Q8y1m4 ralstonia s
38	91.5	7.4	888	16	Q9RD32	Q9rd32 streptomyce
39	91	7.4	257	2	Q9L9U7	Q919u7 chromatium
40	91	7.4	414	16	Q8XUQ2	Q8xuq2 ralstonia s
41	90.5	7.4	324	10	Q8S7Q4	Q8s7q4 oryza sativ
42	90.5	7.4	400	16	Q9RSZ8	Q9rsz8 deinococcus
43	90.5	7.4	683	16	Q8XU96	Q8xu96 ralstonia s
44	89.5	7.3	168	16	Q8XZP1	Q8xzp1 ralstonia s
45	89.5	7.3	209	16	Q9PIA9	Q9pia9 campylobact

ALIGNMENTS

RESULT 1
Q9PCI4
ID Q9PCI4 PRELIMINARY; PRT; 242 AA.
AC Q9PCI4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Xf1795.
GN XF1795.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,

```
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004001; AAF84603.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   242 AA;  25658 MW;  EDC5DE7C0449C1E9 CRC64;

  Query Match           100.0%;  Score 1230;  DB 16;  Length 242;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-93;
  Matches  242;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60

Qy         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120

Qy        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180

Qy        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240

Qy        241 NV 242
              ||
Db        241 NV 242


RESULT 2
Q8Y2M4
ID   Q8Y2M4      PRELIMINARY;      PRT;   295 AA.
AC   Q8Y2M4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcription regulation accessory factor transcription
DE   regulator protein.
GN   RSC0311 OR RS03278.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13839.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Complete proteome.
SQ   SEQUENCE   295 AA;  30990 MW;  133074A7764BFCBB CRC64;

  Query Match           17.4%;  Score 214.5;  DB 16;  Length 295;
  Best Local Similarity  28.4%;  Pred. No. 8.2e-10;
  Matches   78;  Conservative   41;  Mismatches  105;  Indels   51;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVA---------- 50
              | | ||:| | |     | | | | |    | |    | ::|
Db         15 LIDAGNTRIKWAWTAADV-APPAVAPGGTPWQHAGA---RPHDQLAELVEDWRDCHAGAG 70

Qy         51 ------YLASVAAPAITTHVLEVLKIHFE--QVQVAATVAACAGVRIAYAHPERFGVDRF 102
                    ::: || ||:  :  :  |: :::: |: || ||:|  |  | : | ||:
Db         71 MAPPDVWISVVAGPALRDALCARIARVFDGARLRIVASEAAAAGLRNGYRDPAQLGTDRW 130

Qy        103 LALLG---SYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPL 159
              :  :|   :: :  :|:|  ||| |:|::| :|   || |    ||| :||     |||
Db        131 VGAVGARHAWPDTALLLVTAGTATTLDIVAPDGRFAGGLILPGLTLMMRALSRNTAQLPE 190

Qy        160 SGGNYL------------EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLL 206
                ||              :|::|:||:  ||  |    | ::      :     | :|
Db        191 IDIGYLAARDDAQAPADVPSWADNTQDAIALGCVTAQAGAIAQTWQALQAQYPGPYRCVL 250

Qy        207 HGGGVASLLPWLGDVVHRP-----TLVLDGLAIWA 236
               ||  |:| | |      |      ||| || : |
Db        251 SGGARAALAPHL----RMPFQMHDNLVLLGLQVLA 281


RESULT 3
Q9JXF1
ID   Q9JXF1      PRELIMINARY;      PRT;   592 AA.
AC   Q9JXF1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein/Bvg accessory factor.
GN   NMB2075.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002557; AAF42394.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; NMB2075; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  64701 MW;  9E27421DA2B41DE3 CRC64;

  Query Match           16.3%;  Score 200;  DB 16;  Length 592;
  Best Local Similarity  30.6%;  Pred. No. 3.1e-08;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :      |   | |  :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
```

```
                    | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| ::||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
                  |     |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571

Qy        231 GLAIW 235
                | |:
Db        572 NLVIY 576


RESULT 4
Q9JWI7
ID   Q9JWI7      PRELIMINARY;      PRT;   592 AA.
AC   Q9JWI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein (EC 6.3.4.15).
GN   BIRA OR NMA0357.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162752; CAB83659.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   592 AA;  64468 MW;  7ED2E0CD3B31C630 CRC64;

  Query Match             16.1%;  Score 198.5;  DB 16;  Length 592;
  Best Local Similarity   30.7%;  Pred. No. 4.1e-08;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
```

```
                  |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
                : ||
Db        575 IHGL 578


RESULT 5
Q9HWC1
ID   Q9HWC1      PRELIMINARY;      PRT;   248 AA.
AC   Q9HWC1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA4279.
GN   PA4279.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07667.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001230; Prenyl_site.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   248 AA;  26758 MW;  609D37480899DA3B CRC64;

  Query Match             15.7%;  Score 193;  DB 16;  Length 248;
  Best Local Similarity   31.0%;  Pred. No. 3.9e-08;
  Matches   76;  Conservative   34;  Mismatches  111;  Indels   24;  Gaps    8;

Qy          7 DLGNSRFKCASLREGVIGPVTVLPYLTETMDAF-----ALQELPRGRVAYLASVAAPAIT 61
              | |||  |   : ||     :|   | |: ||      : | ||  |   | || :  |
Db          6 DCGNSLIKWRVI-EGAAR--SVAGGLAESDDALVEQLTSQQALP-VRACRLVSVRSEQET 61

Qy         62 THVLEVLKIHFEQVQVAATVAA----CAGVRIAYAHPERFGVDRFLALLGSY--GEGNVL 115
              : ::  |:  |    |:| ||:     ||||  |   :| |:||:|||: ::   :  |
Db         62 SQLVARLEQLF---PVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACL 118

Qy        116 VVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN------YLEFAE 169
              |: :|||:| ||:||:| |||| |    ||||  |     ::             |:  :
Db        119 VIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQ 178

Qy        170 DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVL 229
              | :|:  ||       :     | : |     : | ||    :   |      | ||
Db        179 ATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVF 238

Qy        230 DGLAI 234
                |||:
Db        239 VGLAL 243


RESULT 6
Q9A6Z1
ID   Q9A6Z1      PRELIMINARY;      PRT;   261 AA.
AC   Q9A6Z1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcriptional activator, putative, Baf family.
GN   CC1935.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005867; AAK23910.1; -.
DR   TIGR; CC1935; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   261 AA;  27965 MW;  C19E60D7B0714EF5 CRC64;

  Query Match           11.4%;  Score 140.5;  DB 16;  Length 261;
  Best Local Similarity  27.2%;  Pred. No. 0.00087;
  Matches   49;  Conservative   30;  Mismatches   92;  Indels    9;  Gaps    4;

Qy         50 AYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS- 108
              | : |   |    ::  : : :|    :      | |: :    |   | || :  :|:
Db         60 AVIISSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAA 119

Qy        109 -YGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP-----LSGG 162
               | ::|:  ||| | |::||:|   || |:    |  ||||  | :||      |
Db        120 MVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAG 179

Qy        163 NYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV 222
              | :    ||  |: ||     ::||| :        : : ::   ||||||     | :
Db        180 NRI-VGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIA-TGGVASLFEGATDSI 237


RESULT 7
Q9X8N6
ID   Q9X8N6      PRELIMINARY;      PRT;   265 AA.
AC   Q9X8N6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO3380.
GN   SCO3380 OR SCE94.31C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL049628; CAB40880.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   265 AA;  28203 MW;  E34FA2ABBA946969 CRC64;

  Query Match           11.3%;  Score 139;  DB 16;  Length 265;
  Best Local Similarity  26.5%;  Pred. No. 0.0012;
  Matches   62;  Conservative   33;  Mismatches  107;  Indels   32;  Gaps    9;

Qy         16 ASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQV 75
              | | :|::|          |      ||  |          |::   : || : :: |
Db         40 AVLLQGLMG----------MHPLLGDELGDGIDGIAICATVPSVLHELREVTRRYYGDV 88

Qy         76 -QVAATVAACAGVRIAYAHPERFGVDRFL---ALLGSYGEGNVLVVGVGTALTIDLLAAN 131
               |       || |   ||:  | || :   | :  || |  :||  ||| | | ::|
Db         89 PAVLVEPGVKTGVPILTDHPKEVGADRIINAVAAVELYG-GPAIVVDFGTATTFDAVSAR 147

Qy        132 GCHLGGRISASPTLMRQALHARAEQL-------PLS--GGNYLEFAEDTEDALVSGCNGA 182
              | ::|| |:    :  :||  :  ||       | |  | | :   |  :  :| |  |
Db        148 GEYIGGVIAPGIEISVEALGVKGAQLRKIEVARPRSVIGKNTV---EAMQSGIVYGFAGQ 204

Qy        183 AVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV--HRPTLVLDGLAI 234
              :: |   |      |      ::  ||:| ::     |:  | | | | || :
Db        205 VDGVVNRMARELADDPDDVT--VIATGGLAPMVLGESSVIDEHEPWLTLMGLRL 256


RESULT 8
Q97EB4
ID   Q97EB4      PRELIMINARY;      PRT;   273 AA.
AC   Q97EB4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Predicted transcriptional regulator, homolog of Bvg accessory
DE   factor.
GN   CAC3200.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007815; AAK81136.1; -.
```

```
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30331 MW;  DE7B4D5923E72842 CRC64;

  Query Match             10.6%;  Score 130;  DB 16;  Length 273;
  Best Local Similarity   24.5%;  Pred. No. 0.0067;
  Matches   53;  Conservative   43;  Mismatches  106;  Indels   14;  Gaps    7;

Qy         31 YLTETMDAFALQEL-PRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRI 89
              |  : |: |   :| |       : |   | |   :  ::: :|:   :         |: |
Db         50 YGIQVMNLFQQDKLDPTLVEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINI 109

Qy         90 AYAHPERFGVDRFLALLGSYG--EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMR 147
              | :|:  | || :  : ::   : :::::  ||| |   :  || :||| |     :
Db        110 KYDNPKEVGADRIVNAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSS 169

Qy        148 QALHARAEQLP---LSGGNYL---EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQS 201
             :||  :| :||   |    |          :  :| |  |    ::|| : |  |   :
Db        170 EALFEKAAKLPRVELIKPAYAICKNTISSIQSGIVYGYIGQVRYIVER-MKEELQEEGEK 228

Qy        202 VRLLLHGGGVASLLPWLG---DVVHRPTLVLDGLAI 234
               |::  ||:| |:        ||:: | | |:|| |
Db        229 EPLVVATGGLAKLISEEAKNVDVIN-PFLTLEGLRI 263


RESULT 9
Q9RX54
ID   Q9RX54      PRELIMINARY;      PRT;   262 AA.
AC   Q9RX54;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein DR0461.
GN   DR0461.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001905; AAF10040.1; -.
DR   TIGR; DR0461; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001220; Lectin_legB.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   262 AA;  27839 MW;  965EAAD2F78785A0 CRC64;

  Query Match             10.4%;  Score 128.5;  DB 16;  Length 262;
  Best Local Similarity   27.4%;  Pred. No. 0.0085;
  Matches   58;  Conservative   25;  Mismatches  110;  Indels   19;  Gaps    6;

Qy         39 FALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFG 98
              | |   |  | | |:||| |    : |  || ||     | :      | :    |   |
Db         52 FTLAGAPIPRAAVLSSVAPPVGENYAL-ALKRHFMIDAFAVSAENLPDVTVELDTPGSVG 110

Qy         99 VDRFLALLGSY----GEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARA 154
              ||   | |:      |    :||  ||:   |::      ||| ::    :  || |||
Db        111 ADRLCNLFGAEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARA 170

Qy        155 EQLPLSGGNYLEFA--EDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGG--- 209
             :||       | |  ::|  || ||       ::: |      |      :  ||
Db        171 AKLPRITLQAPETAIGKNTVHALQSGLVFGYAEMVDGLLRRIRAELPGEAVAVATGGFSR 230

Qy        210 ---GVASLLPWLGDVVHRPTLVLDGLA-IWAA 237
                |:   : :  :     || | ||  :||:
Db        231 TVQGICQEIDYYDE-----TLTLRGLVELWAS 257


RESULT 10
Q8XHL5
ID   Q8XHL5      PRELIMINARY;      PRT;   259 AA.
AC   Q8XHL5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CPE2468.
GN   CPE2468.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003194; BAB82174.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28819 MW;  0D5FEA3B7A145E10 CRC64;

  Query Match             10.0%;  Score 122.5;  DB 16;  Length 259;
  Best Local Similarity   22.4%;  Pred. No. 0.026;
  Matches   57;  Conservative   41;  Mismatches  126;  Indels   31;  Gaps    7;

Qy          5 LFDLGNSRF---------KCASLREGVIGPVTVLPYLTETMDAFALQEL-PRGRVAYLAS 54
              | |:||:             || |       |   |  : |  |   :| |      : |
Db          4 LIDVGNTNIVLGIHDNEKYIASWRISTDSKKTSDEYSIQVMQLFNQAKLNPEDVEGIIIS 63

Qy         55 VAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGN- 113
               | |   :  :::  | :  :         |: | | :|:  | || :  : :: :
Db         64 SVVPNIMHSLENMVRKCFCKEPIVVGPGIKTGINIKYDNPKEVGADRIVNAVAAFEKHKK 123

Qy        114 -VLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAE--- 169
              ::::  ||| |   :   | :||| |     :   ||  || :||      :|  :
Db        124 PMIIIDFGTATTFCAITEKGDYLGGNICPGIQISADALFERAAKLP-----RIELEKPKS 178

Qy        170 --------DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDV 221
                     :  :: |  |    :::|    |     | :    :|  ||:| |:    ||
Db        179 VICKNTVTSMQAGIIYGYIGKVEYIVKRMKKEM-MDLGEKEPFVLATGGLAKLVYSETDV 237

Qy        222 VHR--PTLVLDGLAI 234
              :      | |:|| |
Db        238 IDEVDRKLTLEGLKI 252


RESULT 11
Q8R7M2
ID   Q8R7M2      PRELIMINARY;      PRT;   255 AA.
```

```
AC   Q8R7M2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcriptional regulator, homologs of Bvg accessory
DE   factor.
GN   TTE2381.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013180; AAM25520.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   255 AA;  27816 MW;  C3C620ECBC8CA6ED CRC64;

  Query Match             9.7%;  Score 119;  DB 16;  Length 255;
  Best Local Similarity   23.7%;  Pred. No. 0.05;
  Matches   40;  Conservative   33;  Mismatches   76;  Indels   20;  Gaps    6;

Qy          6 FDLGNSRFKCASLR-EGVIGPVTVLPYLTETMDAFALQELPRGRVAY----------L 52
              ||:||:       : : ::    :     :| | : :  |    : |          :
Db          5 FDVGNTNIVMGVFKGKKLLHSFRISTDKNKTYDEYGM--LVNQLIGYNGISLTEIDDVII 62

Qy         53 ASVAAPAITTHVLEVLKI-HFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYG- 110
              :|| | : |  |:|: : :|    :        |: | | :|:  | || :  : :|
Db         63 SSVVPPLMNT--LQVMSLKYFRTKPIVVGPGIKTGINIKYDNPKEVGADRIVNAVAAYEL 120

Qy        111 -EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP 158
               | |:|:  ||| |   ::  | :||| |:    :   ||  |  :||
Db        121 YGGPVIVIDFGTATTFCAISEKGEYLGGIIAPGLMISADALFQRTAKLP 169


RESULT 12
Q9F985
ID   Q9F985      PRELIMINARY;      PRT;   258 AA.
AC   Q9F985;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative 32 kDa replication protein.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=V;
RA   Vasquez C., Pichuantes S., Saavedra C.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF198621; AAG28531.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
SQ   SEQUENCE   258 AA;  28101 MW;  507B55D695095855 CRC64;

  Query Match             9.5%;  Score 117;  DB 2;  Length 258;
  Best Local Similarity   27.4%;  Pred. No. 0.074;
  Matches   54;  Conservative   36;  Mismatches   83;  Indels   24;  Gaps    9;

Qy         52 LASVAAPAITTHVLEVLK-IHFEQVQVAATVAACAGVRIAYAHPERFGVDRFL-ALLGSY 109
              ::||  | :       ||  | : : |    :    |: | | :|   | || : |: | :
Db         62 ISSVVPPIMFALERMCLKYFHIKPIIVGPGIK--TGLNIKYDNPREVGADRIVNAVAGIH 119

Qy        110 GEGN-VLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFA 168
               |: :::|  ||| |   :  :  ::|| |:    :  :|| ||| :||      :| |
Db        120 LYGSPLIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLP-----RIEIA 174

Qy        169 -------EDTEDALVSGCNGAAVALIER--SLYEAHQRLDQSVRLLLHGGGVASLLPWLG 219
                    ::|  |: :|     |  :|   |  :|   :   |   :  ||:|||:
Db        175 RPDDIIGKNTVSAMQAGILYGYVGQVEGIVSRMKAKSPVPPKV---IATGGLASLIASES 231

Qy        220 DVVH--RPTLVLDGLAI 234
              :|:   | | | || |
Db        232 NVIDIVDPFLTLTGLKI 248


RESULT 13
O25533
ID   O25533      PRELIMINARY;      PRT;   223 AA.
AC   O25533;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein HP0862.
GN   HP0862.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000596; AAD07916.1; -.
DR   TIGR; HP0862; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   223 AA;  24655 MW;  C1D13E8BEC1EB9AF CRC64;

  Query Match             9.4%;  Score 115.5;  DB 16;  Length 223;
  Best Local Similarity   31.5%;  Pred. No. 0.082;
  Matches   45;  Conservative   22;  Mismatches   55;  Indels   21;  Gaps    6;

Qy         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152
              |:|| :| |       | :||  |:|:||||:   | |||| |      : | :|     |
Db         85 GIDRQMACL---AVNNGVVVDAGSAITIDLI-KEGKHLGGCILPG---LAQYIHAYKKSA 137

Qy        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211
              :  : |    : ||   : | ||:  |   : :| |        | | :: :: | ||
Db        138 KILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACI--------QHLAKNQKIYLCGGDA 189

Qy        212 ASLLPWLGDVVHRPTLVLDGLAI 234
                |  :|   | :  || ||: |
Db        190 KYLSAFLPHSVCKERLVFDGMEI 212


RESULT 14
Q9CD56
ID   Q9CD56      PRELIMINARY;      PRT;   274 AA.
AC   Q9CD56;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
```

```
DE   Hypothetical protein ML0232.
GN   ML0232.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; AL583917; CAC29740.1; -.
DR   Leproma; ML0232; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   274 AA;  29421 MW;  1C2E735BDEC78765 CRC64;

  Query Match             9.2%;  Score 113.5;  DB 16;  Length 274;
  Best Local Similarity   26.5%;  Pred. No. 0.15;
  Matches   54;  Conservative   30;  Mismatches   95;  Indels   25;  Gaps    8;

Qy         52 LASVAAPAITTHVLEVLKIHFEQV-----QVAATVAACAGVRIAYAHPERFGVDRF---L 103
              ||  || :    ||  ::|  :|       |       |: :   :|:  | ||    |
Db         58 LAGAAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        104 ALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQL------ 157
              |    :|:   :||  |::: :|:::| |  ||| |:    :   |  ||:  |
Db        118 AAFHKFGQA-AIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELA 176

Qy        158 -PLS--GGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQ---RLDQSVRLLLHGGGV 211
               | |  | | :|     :  :| |  |    |: |   :  :    |   | ::  |
Db        177 RPRSVVGKNTVEC---MQAGVVFGFAGLVDGLVGRMRQDVEEFSGDLGNRVAVVATGHTA 233

Qy        212 ASLLPWLGDVVHRPT-LVLDGLAI 234
                |||  |  |    | | || :
Db        234 PLLLPELHTVDHYDRHLTLHGLRL 257


RESULT 15
O06282
ID   O06282      PRELIMINARY;      PRT;   272 AA.
AC   O06282;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 29.3 kDa protein (Transcriptional activator, putative,
DE   Baf family).
GN   RV3600C OR MTCY07H7B.22 OR MT3706.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z95557; CAB08944.1; -.
DR   EMBL; AE007170; AAK48063.1; -.
DR   TIGR; MT3706; -.
DR   TubercuList; Rv3600c; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   272 AA;  29304 MW;  5D70E6EF0F09AC8B CRC64;

  Query Match             9.1%;  Score 112;  DB 16;  Length 272;
  Best Local Similarity   25.3%;  Pred. No. 0.2;
  Matches   56;  Conservative   32;  Mismatches  107;  Indels   26;  Gaps    8;

Qy         32 LTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQV-----QVAATVAACAG 86
              |  |:|    ::  |     |   || :    ||  ::|  :|       |       |
Db         43 LALTIDGLIGEDSER-----LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTG 97

Qy         87 VRIAYAHPERFGVDRFLALLGSYGEGN--VLVVGVGTALTIDLLAANGCHLGGRISASPT 144
              : :   :|:  | || :  | :|       :||  |::: :|:::| |  ||| |:
Db         98 IPLLVDNPKEVGADRIVNCLAAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQ 157

Qy        145 LMRQALHARAEQL-------PLS--GGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAH 195
              :   |  ||:  |       | |  | | :|     :   | |  |    |: |   :
Db        158 VSSDAAAARSAALRRVELARPRSVVGKNTVEC---MQAGAVFGFAGLVDGLVGRIREDVS 214

Qy        196 Q-RLDQSVRLLLHGGGVASLLPWLGDVVHRPT-LVLDGLAI 234
                :|  | ::  |     ||| |  | |     | | || :
Db        215 GFSVDHDVAIVATGHTAPLLLPELHTVDHYDQHLTLQGLRL 255


Search completed: June 24, 2003, 22:16:24
Job time : 23.2874 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:51:40 ; Search time 4.86805 Seconds
(without alignments)
2061.866 Million cell updates/sec

Title: US-09-813-453A-65
Perfect score: 1230
Sequence: 1 MNDWLFDLGNSRFKCASLRE.....HRPTLVLDGLAIWAAVAANV 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	194.5	15.8	267	1	BAF_BORPE	Q45338 bordetella
2	116	9.4	233	1	YACB_BACSU	P37564 bacillus su
3	94.5	7.7	655	1	DNK1_SYNP7	P50020 synechococc
4	93	7.6	399	1	HUTI_DEIRA	Q9rz05 deinococcus
5	92.5	7.5	397	1	NCCB_ALCXX	Q44585 alcaligenes
6	91	7.4	407	1	MURG_MYCLE	O69552 mycobacteri
7	90.5	7.4	260	1	Y230_ARCFU	O30009 archaeoglob
8	90	7.3	301	1	PPNK_STRCO	Q9s219 streptomyce
9	89.5	7.3	3491	1	ERY1_SACER	Q03131 saccharopol
10	88.5	7.2	255	1	GLO2_RHOBL	P05446 rhodopseudo
11	88.5	7.2	488	1	DHAL_BACST	P42329 bacillus st
12	88	7.2	1609	1	CTPI_MYCLE	O53114 mycobacteri
13	87.5	7.1	381	1	GRK1_ECOLI	P77364 escherichia
14	86.5	7.0	1569	1	YPJA_ECOLI	P52143 escherichia
15	85.5	7.0	920	1	MML7_MYCTU	P96289 mycobacteri
16	85	6.9	534	1	LY41_STRCO	O86511 streptomyce
17	84.5	6.9	381	1	TAL_STRCO	O88018 streptomyce
18	84	6.8	540	1	FLIF_BRUAB	O52069 brucella ab
19	83.5	6.8	668	1	BIP5_TOBAC	Q03685 nicotiana t
20	83	6.7	563	1	4CL1_ORYSA	P17814 oryza sativ
21	83	6.7	1083	1	EMBB_MYCLE	Q9cda9 mycobacteri
22	82.5	6.7	656	1	HS7C_DROME	P29844 drosophila
23	82.5	6.7	666	1	BIP_LYCES	P49118 lycopersico
24	82	6.7	362	1	HUPK_BRAJA	P48342 bradyrhizob
25	82	6.7	409	1	PROA_MYCLE	Q9cbz7 mycobacteri
26	82	6.7	1065	1	EMBB_MYCAV	P71486 mycobacteri
27	82	6.7	1094	1	EX5B_MYCTU	P96920 mycobacteri
28	81.5	6.6	214	1	CCMA_RHOCA	P29959 rhodobacter
29	81.5	6.6	278	1	Y4LA_RHISN	P55541 rhizobium s
30	81.5	6.6	344	1	YD33_MYCTU	Q10644 mycobacteri
31	81.5	6.6	654	1	GR78_HUMAN	P11021 homo sapien
32	81.5	6.6	654	1	GR78_MESAU	P07823 mesocricetu
33	81.5	6.6	654	1	GR78_RAT	P06761 rattus norv
34	81.5	6.6	655	1	GR78_MOUSE	P20029 mus musculu
35	81.5	6.6	668	1	BIP2_ARATH	Q39043 arabidopsis
36	81.5	6.6	669	1	BIP1_ARATH	Q9lkr3 arabidopsis
37	81	6.6	483	1	CBF5_YEAST	P33322 saccharomyc
38	81	6.6	1085	1	CARB_HELPJ	Q9zkt2 helicobacte
39	80.5	6.5	323	1	ISPB_ECOLI	P19641 escherichia
40	80.5	6.5	488	1	DHA1_BACSU	P42236 bacillus su
41	80.5	6.5	652	1	GR78_CHICK	Q90593 gallus gall
42	80.5	6.5	667	1	BIP4_TOBAC	Q03684 nicotiana t
43	80.5	6.5	786	1	AAC_ACTUT	P29958 actinoplane
44	80.5	6.5	811	1	MUTS_THEAQ	Q56215 thermus aqu
45	80	6.5	439	1	GUAD_DEIRA	Q9ryx4 deinococcus

ALIGNMENTS

RESULT 1
BAF_BORPE
ID BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshazer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli.";
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN=BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a birA homolog in Bordetella pertussis.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN=BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus
RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis.";
RL Mol. Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U12020; AAA75361.1; -.
DR EMBL; AF016461; AAC68834.1; -.
DR EMBL; X90711; CAA62242.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

```
  Query Match             15.8%;  Score 194.5;  DB 1;  Length 267;
  Best Local Similarity   27.9%;  Pred. No. 2.5e-09;
  Matches   80;  Conservative   25;  Mismatches  105;  Indels   77;  Gaps   10;

Qy          5 LFDLGNSRFKCA--------SLREGVIGPVTVLPYLTETMDAFA----LQELPRG----- 47
              | | |||| |           : ||         |   : :|  |    |  |||
Db          4 LIDSGNSRLKVGWFDPDAPQAAREPA-------PVAFDNLDLDALGRWLATLPRRPQRAL 56

Qy         48 --RVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAA------------CAGVRIAYA 92
                 || ||   |                    :|||: |              |:|  |
Db         57 GVNVAGLARGEA------------------IAATLRAGGCDIRWLRAQPLAMGLRNGYR 97

Qy         93 HPERFGVDRFLALLGSYG-----EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMR 147
              :|:: | ||:  ::|          :||   ||| |:| :  :    || |   | :||
Db         98 NPDQLGADRWACMVGVLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMR 157

Qy        148 QALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLH 207
               ||      |||: |   ::  ||  |: ||   |    | |       ||  |:  : :
Db        158 GALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQAGAIVRQWLAGRQRYGQAPEIYVA 217

Qy        208 GGG-------VASLLPWLG---DVVHRPTL----VLDGLAIWAAVAA 240
              |||          ||   |       :||     ||||||  ||  |
Db        218 GGGWPEVRQEAERLLAVTGAAFGATPQPTYLDSPVLDGLAALAAQGA 264
```

```
RESULT 2
YACB_BACSU
ID   YACB_BACSU     STANDARD;      PRT;   233 AA.
AC   P37564;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yacB.
GN   YACB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             9.4%;  Score 116;  DB 1;  Length 233;
  Best Local Similarity   23.0%;  Pred. No. 0.0081;
  Matches   47;  Conservative   34;  Mismatches   75;  Indels   48;  Gaps    7;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHV 64
              |||      |:     :|:|    | | :      |||: :           :  | : |
Db         46 LFDHSGLMFEQI---DGIIISSVVPPIM------FALERMCTKYFHIEPQIVGPGMKT-- 94

Qy         65 LEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRF---LALLGSYGEGNVLVVGVGT 121
                                    |: | | :|:  | ||    :| :  ||   ::||  ||
Db         95 --------------------GLNIKYDNPKEVGADRIVNAVAAIHLYG-NPLIVVDFGT 132

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP---------LSGGNYLEFAEDTE 172
               | |   :  |  ::|| |:   |:  :||::|| :||         : | | :      :
Db        133 ATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNTV---SAMQ 189

Qy        173 DALVSGCNGAAVALIERSLYEAHQ 196
              :: |  |    :::|  ::| |
Db        190 SGILFGYVGQVEGIVKRMKWQAKQ 213
```

```
RESULT 3
DNK1_SYNP7
ID   DNK1_SYNP7     STANDARD;      PRT;   655 AA.
AC   P50020;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Chaperone protein dnaK1 (Heat shock protein 70-1) (Heat shock 70 kDa
DE   protein 1) (HSP70-1).
GN   DNAK1.
OS   Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1140;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94257019; PubMed=8198610;
RA   Nimura K., Yoshikawa H., Takahashi H.;
RT   "Identification of dnaK multigene family in Synechococcus sp.
RT   PCC7942.";
RL   Biochem. Biophys. Res. Commun. 201:466-471(1994).
CC   -!- FUNCTION: ACTS AS A CHAPERONE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D28550; BAA05903.1; -.
DR   HSSP; P04475; 1DG4.
DR   InterPro; IPR001023; Hsp70.
DR   Pfam; PF00012; HSP70; 1.
DR   PRINTS; PR00301; HEATSHOCK70.
DR   ProDom; PD000089; Hsp70; 1.
DR   PROSITE; PS00297; HSP70_1; 1.
DR   PROSITE; PS00329; HSP70_2; 1.
DR   PROSITE; PS01036; HSP70_3; 1.
KW   Chaperone; ATP-binding; Heat shock; Phosphorylation; Multigene family.
FT   MOD_RES     197    197       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ   SEQUENCE   655 AA;  71058 MW;  AE65604C3E70A9DA CRC64;

  Query Match             7.7%;  Score 94.5;  DB 1;  Length 655;
  Best Local Similarity   20.8%;  Pred. No. 1.6;
  Matches   59;  Conservative   44;  Mismatches   93;  Indels   87;  Gaps   13;

Qy         34 ETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAAT--VAACAGVRIAY 91
              : : |  ||:| :    :|      |: |     :  :|   |  ||    | ||: :
Db        111 QEISAMVLQKLRQDAETFLGEPVTQAVIT-----VPAYFTDAQRQATKDAGAIAGLEVLR 165

Qy         92 AHPE------RFGVDRFLALLGSYGEGNVLVVGV-GTALTIDLL-----------AANGC 133
                |      :|:|:       :    :||  : |: | : :|             |
Db        166 IVNEPTAAALSYGLDKL------HENSRILVFDLGGGSTLDVSILQLGDSVFEVKATAGNN 219

Qy        134 HLGGR--------------ISASPTLMR------QALHARAEQ-------LPLSGGNY-- 164
              ||||              : |    :|      | |   :||       || :  |
Db        220 HLGGDDFDAVIVDWLADNFLKAESIDLRQDKMAIQRLREASEQAKIDLSTLPTTTINLPF 279

Qy        165 ----------------LEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHG 208
                            :|   :  : | |    | :  |:::| :::  :||  |:|| |
Db        280 IATATVDGAPEPKHIEVELQREQFEVLASNLVQATIEPIQQALKDSNLTIDQIDRILLVG 339

Qy        209 GGVASLLPWLGDVVHR------PTLVL---DGLAIWAAVAANV 242
              |  :| :| :   | :      | | :   : :|: ||: | |
Db        340 G--SSRIPAIQQAVQKFFGGKTPDLTINPDEAIALGAAIQAGV 380


RESULT 4
HUTI_DEIRA
ID   HUTI_DEIRA     STANDARD;      PRT;   399 AA.
AC   Q9RZ05;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Imidazolonepropionase (EC 3.5.2.7) (Imidazolone-5-propionate
DE   hydrolase).
GN   HUTI OR DRA0148.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- CATALYTIC ACTIVITY: 4-imidazolone-5-propanoate + H(2)O = N-
CC       formimino-L-glutamate.
CC   -!- PATHWAY: Histidine degradation; third step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO THE HUTI FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE001862; AAF12215.1; -.
DR   TIGR; DRA0148; -.
DR   InterPro; IPR002604; ATZ_TRZ.
DR   Pfam; PF01685; ATZ_TRZ; 1.
DR   TIGRFAMs; TIGR01224; hutI; 1.
KW   Hydrolase; Histidine metabolism; Complete proteome.
SQ   SEQUENCE   399 AA;  41572 MW;  0D8A66747AB79D11 CRC64;

  Query Match             7.6%;  Score 93;  DB 1;  Length 399;
  Best Local Similarity   27.6%;  Pred. No. 1.3;
  Matches   64;  Conservative   19;  Mismatches   87;  Indels   62;  Gaps   13;

Qy         15 CASL-----REGVIGPVTVLPYLTETMDAFALQE---LPRGRVAYLASVAAPAITTHVLE 66
              || |     |||          | | :| |  :|    :   |  : |: |      | |
Db        191 CAELIPQVAREG----------LAEALDVFCEKEAFSVEETRTMFAAAQA------HGLR 234

Qy         67 VLKIHFEQVQ-VAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVGTALTI 125
              | |:| :|   :  |  ||    ::  | |  |  :  ||  |     :|    | | :
Db        235 V-KLHADQPHAIGGTELACEVGALSVDHLEASGAAQIAALAASETVATIL---PGVTLHL 290

Qy        126 DLLAANGCHL--GGRISA---------SPTLMRQALHARAEQLPLSGGNYLEFAEDTEDA 174
                | || |  |   | | |          ||     |   | | :| |            |
Db        291 GLPAAPGRQLIDSGAIVAIGTDLNPGSSPLFSTQLALALAVRLCLL----------TPAE 340

Qy        175 LVSGC--NGA-AVALIERSLYEAHQRLD---------QSVRLLLHGGGVASL 214
                :| |  | | |: | :|    | || |          : |   | |  || :
Db        341 ALSACTVNAAYALGLSDRGSLSAGQRADFLVLNGQDWREVAYTLGGNAVAEV 392


RESULT 5
NCCB_ALCXX
ID   NCCB_ALCXX     STANDARD;      PRT;   397 AA.
AC   Q44585;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Nickel-cobalt-cadmium resistance protein nccB.
GN   NCCB.
OS   Alcaligenes xylosoxydans xylosoxydans (Achromobacter xylosoxidans).
OG   Plasmid pTOM9.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Achromobacter.
OX   NCBI_TaxID=515;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=31A;
RX   MEDLINE=95050278; PubMed=7961470;
RA   Schmidt T., Schlegel H.G.;
RT   "Combined nickel-cobalt-cadmium resistance encoded by the ncc locus
RT   of Alcaligenes xylosoxidans 31A.";
RL   J. Bacteriol. 176:7045-7054(1994).
CC   -!- FUNCTION: COMPONENT OF THE NCC CATION-EFFLUX SYSTEM THAT CONFERS
CC       RESISTANCE TO NICKEL, COBALT AND CADMIUM.
CC   -!- SIMILARITY: HIGH, TO A.EUTROPHUS CNRB AND SOME, TO ALCALIGENES
CC       CZCB.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L31363; AAA65105.1; -.
DR   InterPro; IPR002215; HlyD.
DR   Pfam; PF00529; HlyD; 1.
DR   TIGRFAMs; TIGR00999; 8a0102; 1.
KW   Plasmid; Nickel; Cobalt; Cadmium; Transport.
SQ   SEQUENCE   397 AA;  39933 MW;  C6DE57CE314996B8 CRC64;

  Query Match             7.5%;  Score 92.5;  DB 1;  Length 397;
  Best Local Similarity   25.2%;  Pred. No. 1.4;
  Matches   55;  Conservative   23;  Mismatches   73;  Indels   67;  Gaps    9;

Qy         42 QELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACA-----GVRIAYAHPER 96
              ||:   |:|                    | :   :|| |||||  |     |  :|   |
Db        174 QEMESARIA----------------LDVAQAEVQRAATVAQAAKVSSDGRSVAVVSPIA 216

Qy         97 FGVDRFLALLGSY-----------GEGNVLV-----------VGVGTALTIDLLAANGCH 134
                :      ||:|           | | | |           :  |:  || |  |||
Db        217 GRITAQSVTLGAYVAPQAELFRVAGSGAVQVEAYVTAADTSRIAAGSDATIVL--ANGAP 274

Qy        135 LGGRISA-SPTLMRQALHARAEQLPLSGGNYLEFAEDTEDAL-VSGCNGAAVALIERSLY 192
              | ||: | :||:   |  |     |:   : |   |  :  |     :  |::: |
Db        275 LAGRVQAVTPTVSGSARAATVVVTPVDANSGLIVGEGVQVRLHTKAADANAMSVPE---- 330

Qy        193 EAHQRLD--------------QSVRLLLHGGGVASLL 215
              :| | ||               :|| :    |||| :|
Db        331 DAVQNLDGRDVVFVRTQQGFRPKSVLVGSRSGGVAQIL 368


RESULT 6
MURG_MYCLE
ID   MURG_MYCLE     STANDARD;      PRT;   407 AA.
AC   O69552;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   UDP-N-acetylglucosamine--N-acetylmuramyl-(Pentapeptide)
DE   pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
DE   (EC 2.4.1.-) (Undecaprenyl-PP-MurNAc-pentapeptide-UDPGlcNAc GlcNAc
DE   transferase).
GN   MURG OR ML0914 OR MLCB268.02C.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
  P  SEQUENCE FROM N.A.
  C  STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- FUNCTION: CELL WALL FORMATION. CATALYZES THE TRANSFER OF A GLCNAC
CC       SUBUNIT ON UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-PENTAPEPTIDE (LIPID
CC       INTERMEDIATE I) TO FORM UNDECAPRENYL-PYROPHOSPHORYL-MURNAC-
CC       (PENTAPEPTIDE)GLCNAC (LIPID INTERMEDIATE II) (BY SIMILARITY).
CC   -!- PATHWAY: Peptidoglycan biosynthesis; last step.
CC   -!- SUBCELLULAR LOCATION: Membrane-associated (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE MURG FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AL022602; CAA18668.1; -.
DR   EMBL; AL583920; CAC31295.1; -.
DR   HSSP; P17443; 1F0K.
DR   Leproma; ML0914; -.
DR   InterPro; IPR004276; Glyco_transf_28.
DR   Pfam; PF03033; Glyco_transf_28; 1.
DR   TIGRFAMs; TIGR01133; murG; 1.
KW   Transferase; Glycosyltransferase; Cell division; Cell wall; Membrane;
KW   Peptidoglycan synthesis; Complete proteome.
SQ   SEQUENCE   407 AA;  42422 MW;  B757B73C092C53F9 CRC64;

  Query Match             7.4%;  Score 91;  DB 1;  Length 407;
  Best Local Similarity   25.1%;  Pred. No. 1.9;
  Matches   58;  Conservative   27;  Mismatches   88;  Indels   58;  Gaps    8;

Qy         25 PVTVLPYLTETMDAFALQE-----------LPRGRVAYLASVAAPAITTHVLEVLKIHF 72
              |: |   : ||   | : |             ||    ||||:   | :   :  |:
Db        102 PLRVWRAVRETRAVFEVVEAHVVVGFGGYVALP----AYLAARGIPRVRRRIPVVV---- 153

Qy         73 EQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANG 132
              :    | :|   ||| |          :| |: :   |     ||||    ||
Db        154 HEANARAGIANRVGVRTA---------ERVLSAVPGSGLRGAEVVGVPIHATI------- 197

Qy        133 CHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLY 192
                        || | || | | :       :  | :|    || | :  ||:|  |::
Db        198 ----------TTLNRPALRADARK-------HFGFTDDARVLLVFGGSQGAVSL-NRAVA 239

Qy        193 EAHQRLDQSVRLLLHGGGVASLL----PWLGDVVHRPTLVLDGLAIWAAVA 239
              | : |  |   :||  |: : |    |  |:  :      || : :  | |
Db        240 GAAEDLAASGVAVLHAYGLKNTLELRTPEYGEPPYVAVPYLDRMDLAYAAA 290


RESULT 7
Y230_ARCFU
ID   Y230_ARCFU     STANDARD;      PRT;   260 AA.
AC   O30009;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative aldolase AF0230 (EC 4.2.1.-).
GN   AF0230.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC   Archaeoglobaceae; Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
CC   -!- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AE001090; AAB91002.1; -.
DR   TIGR; AF0230; -.
DR   InterPro; IPR002915; DeoC.
DR   Pfam; PF01791; DeoC; 1.
KW   Hypothetical protein; Lyase; Schiff base; Complete proteome.
FT   BINDING     172    172       SCHIFF-BASE (BY SIMILARITY).
SQ   SEQUENCE   260 AA;  28196 MW;  3A2FC086D702A2D0 CRC64;

  Query Match             7.4%;  Score 90.5;  DB 1;  Length 260;
  Best Local Similarity   22.6%;  Pred. No. 1.2;
  Matches   40;  Conservative   39;  Mismatches   71;  Indels   27;  Gaps    8;

Qy         54 SVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGN 113
              :: | |:: ||    |   ||:: | |::          : :|:   ||::  |  |:
Db        101 ALGADAVSIHVNIGSKTEAEQIEKAGTISEIC---------DDYGIP-LLAMM--YPRGS 148

Qy        114 VLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPL-----SGGNYLEFA 168
              : |    |  |:   |  |  ||  |   | :           :|:     | |:  ||
Db        149 IDV----TTETVRHAARIGYELGADILKVPYVQSFEEVVAVCDIPVVVAGGSKGSEHEFL 204

Qy        169 EDTEDALVSGCNGAAVALIERSLYEAHQ--RLDQSVRLLLHGG-GVASLLPWLGDVV 222
              :  |||:  |  | |     |::: :    |: ::: :::||   :  :: : |::|
Db        205 KRVEDAIAKGAAGVAAG---RNVFNSDHPVRIAKALHMIVHGNMHMEEVMEYEGNMV 258


RESULT 8
PPNK_STRCO
ID   PPNK_STRCO     STANDARD;      PRT;   301 AA.
AC   Q9S219;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable inorganic polyphosphate/ATP-NAD kinase (EC 2.7.1.23)
DE   (Poly(P)/ATP NAD kinase).
GN   PPNK OR SCO1781 OR SCI51.21C.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- FUNCTION: Catalyzes the phosphorylation of NAD to NADP. Utilizes
CC       ATP and other nucleoside triphosphates as well as inorganic
CC       polyphosphate as a source of phosphorus (By similarity).
CC   -!- CATALYTIC ACTIVITY: ATP + NAD(+) = ADP + NADP(+).
CC   -!- COFACTOR: Requires divalent metal ions for activity (By
CC       similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE NAD KINASE FAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL109848; CAB52845.1; -.
DR   InterPro; IPR002504; ATP_NADK.
DR   Pfam; PF01513; NAD_kinase; 1.
KW   Transferase; Kinase; NAD; NADP; Complete proteome.
SQ   SEQUENCE   301 AA;  32081 MW;  FCB6F6B01766AFBA CRC64;

  Query Match             7.3%;  Score 90;  DB 1;  Length 301;
  Best Local Similarity   27.6%;  Pred. No. 1.6;
  Matches   50;  Conservative   22;  Mismatches   59;  Indels   50;  Gaps   10;

Qy         49 VAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS 108
              |  ||    ||          |   :: |   : |  |||:  |        | | |
Db          9 VFLLAHTGRPA---------AIRSAELVVKGLLRAGIGVRVLEAEA------RDLPL--- 50

Qy        109 YGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN----- 163
              | | :||  |   :|     || |   :    ||:| |  |||  :|: | |
Db         51 --PGEVELVGEATPQCLD-----GCELLIVLGGDGTLLRGAEFARASGVPMLGVNLGRVG 103

Qy        164 YLEFAE-DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV 222
              :|  || |  | :|        :: |: ||  :|:  :| :::|          ||:|
Db        104 FLAEAERDDLDKVVD-------RVVNRA-YEVEERM--TVDVVVHRN---------GDIV 144

Qy        223 H 223
              |
Db        145 H 145


RESULT 9
ERY1_SACER
ID   ERY1_SACER     STANDARD;      PRT;  3491 AA.
AC   Q03131;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 1 and 2 (EC 2.3.1.94) (ORF 1) (6-
DE   deoxyerythronolide B synthase I) (DEBS 1).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [2]
RP   SEQUENCE OF 3474-3491 FROM N.A.
RX   MEDLINE=93231529; PubMed=8386127;
RA   Donadio S., Staver M.J.;
RT   "IS1136, an insertion element in the erythromycin gene cluster of
RT   Saccharopolyspora erythraea.";
RL   Gene 126:147-151(1993).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
```

```
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 3 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M63676; AAA26493.2; -.
DR   EMBL; L07626; AAA26504.1; -.
DR   HSSP; P25715; 1MLA.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 3.
DR   Pfam; PF00698; Acyl_transf; 3.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 3.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1972       MODULE 1.
FT   DOMAIN     1979   3491       MODULE 2.
FT   DOMAIN        1    375       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN      414    484       ACYL CARRIER (ACP) 1.
FT   DOMAIN      503    961       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN     1030   1356       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     1611   1794       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1888   1958       ACYL CARRIER (ACP) 2.
FT   DOMAIN     1979   2441       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2507   2854       ACYLTRANSFERASE (AT) 3.
FT   DOMAIN     3055   3237       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     3334   3404       ACYL CARRIER (ACP) 3.
FT   ACT_SITE    145    145       ACYL-ENZYME INTERMEDIATE.
FT   BINDING     447    447       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE    677    677       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   1128   1128       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1614   1660       NADP.
FT   BINDING    1921   1921       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   2148   2148       BETA-KETOACYL SYNTHASE.
FT   ACT_SITE   2598   2598       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    3058   3104       NADP.
FT   BINDING    3367   3367       PHOSPHOPANTETHEINE (BY SIMILARITY).
SQ   SEQUENCE   3491 AA;  365022 MW;  682BFC32C90FA8C4 CRC64;

  Query Match             7.3%;  Score 89.5;  DB 1;  Length 3491;
  Best Local Similarity   22.7%;  Pred. No. 29;
  Matches   63;  Conservative   30;  Mismatches  105;  Indels   79;  Gaps   12;

Qy          8 LGNSRPKCASLREGV-IGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVL- 65
              :   | :  | |  | || | |    :| :|  |     |  :  :  || |:    :
Db          1 MSGPRSRTTSRRTPVRIGAVVVASSTSELLDGLAAVADGRPHASVVRGVARPSAPVVFVF 60

Qy         66 ------------EVLKIHFEQVQVAATVAACAGV-------------------RIAYAH 93
                        |:|    |    || : |||                      |:
Db         61 PGQGAQWAGMAGELLG---ESRVFAAAMDACARAFEPVTDWTLAQVLDSPEQSRRVEVVQ 117

Qy         94 PERFGVDRFLALL-GSYGEGNVLVVG--VGTALTIDLLAANGCHLGGRISASPTLMRQAL 150
              |  | |   || |  |:|     |||  :|     :| ||: |   |   |:      ||
Db        118 PALFAVQTSLAALWRSFGVTPDAVVGHSIG-----ELAAAHVC---GAAGAADAARAAAL 169

Qy        151 HARAEQLPLSGGNYL----------EFAEDTEDALVSGCNGAAVALIERSLYEAHQRL 198
              :| | :|| |   :              |   :| :::| ||    |:  |     :|:
Db        170 WSR-EMIPLVGNGDMAAVALSADEIEPRIARWDDDVVLAGVNGPRSVLLTGSPEPVARRV 228

Qy        199 DQ-----------SVRLLLHGG-------GVASLLPW 217
                          :|  :  |         |: | | |
Db        229 QELSAEGVRAQVINVSMAAHSAQVDDIAEGMRSALAW 265


RESULT 10
GLO2_RHOBL
ID   GLO2_RHOBL     STANDARD;      PRT;   255 AA.
AC   P05446;
DT   01-NOV-1988 (Rel. 09, Created)
DT   01-NOV-1988 (Rel. 09, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable hydroxyacylglutathione hydrolase (EC 3.1.2.6) (Glyoxalase
DE   II) (Glx II).
GN   GLOB.
OS   Rhodopseudomonas blastica.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1075;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=85058188; PubMed=6209404;
RA   Tybulewicz V.L.J., Falk G., Walker J.E.;
RT   "Rhodopseudomonas blastica atp operon. Nucleotide sequence and
RT   transcription.";
RL   J. Mol. Biol. 179:185-214(1984).
CC   -!- FUNCTION: THIOLESTERASE THAT CATALYSES THE HYDROLYSIS OF S-D-
CC       LACTOYL-GLUTATHIONE TO FORM GLUTATHIONE AND D-LACTIC ACID (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: (S)-(2-hydroxyacyl)glutathione + H(2)O =
CC       glutathione + a 2-hydroxy acid anion.
CC   -!- COFACTOR: BINDS TWO ZINC IONS (BY SIMILARITY).
CC   -!- PATHWAY: Glyoxal pathway.
CC   -!- SIMILARITY: BELONGS TO THE GLYOXALASE II FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z00018; CAA77309.1; -.
DR   PIR; S04668; S04668.
DR   HSSP; Q16775; 1QH5.
DR   InterPro; IPR001279; Blactmase-like.
DR   Pfam; PF00753; lactamase_B; 1.
KW   Hydrolase; Zinc.
FT   METAL        56     56       ZINC 1 (BY SIMILARITY).
FT   METAL        58     58       ZINC 1 (BY SIMILARITY).
FT   METAL        60     60       ZINC 2 (BY SIMILARITY).
FT   METAL        61     61       ZINC 2 (BY SIMILARITY).
FT   METAL       114    114       ZINC 1 (BY SIMILARITY).
FT   METAL       133    133       ZINC 1 AND 2 (BY SIMILARITY).
FT   METAL       171    171       ZINC 2 (BY SIMILARITY).
SQ   SEQUENCE   255 AA;  27200 MW;  62A7573A3F4991DC CRC64;

  Query Match             7.2%;  Score 88.5;  DB 1;  Length 255;
  Best Local Similarity   25.0%;  Pred. No. 1.8;
  Matches   48;  Conservative   25;  Mismatches   70;  Indels   49;  Gaps   10;

Qy         29 LPYLTETMDAFALQELPRGRVAYL-ASVAAPAITT------HVLEVLKIHF--EQVQVAA 79
              :| | :   || | :   || | : |  ||| ::        : ::|  |   : :|
Db          8 VPCLKDNF-AFLLHDAASGRTALVDAPEAAPVLSALAAQGWRLTDILLTHHHDDHIQAVP 66

Qy         80 TVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVGTALTIDL------------ 127
              : |  | |:   |       |   |  :  ||: : :|  :|: ||:
```

```
Db         67 EIHAATGARV---HGAAADAHRLPPLDHALREGDRVAIGAESAVVIDVPGHTRGHIAFHF 123

Qy        128 -----------LAANGCHLGGRISASPTLMRQALHARA----EQLPLSGGNYLE----FA 168
                          | | ||  |     :|  |  :|   |    | |  || :||:    ||
Db        124 PGSALAFTGDSLMAAGC--GRLFEGTPAEMWASLSKLAALPPETLICSGHDYLDGNLRFA 181

Qy        169 ---EDTEDALVS 177
                 |    ||:|
Db        182 LALEPGNPALIS 193


RESULT 11
DHAL_BACST
ID   DHAL_BACST     STANDARD;      PRT;   488 AA.
AC   P42329;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aldehyde dehydrogenase, thermostable (EC 1.2.1.3).
GN   ALDHT.
OS   Bacillus stearothermophilus.
OC   Bacteria; Firmicutes; Bacillales; Geobacillus.
OX   NCBI_TaxID=1422;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SIC1;
RA   Imanaka T., Ohta T., Sakoda H., Widhyastuti N., Matsuoka M.;
RT   "Cloning, nucleotide sequence, and efficient expression of the gene
RT   coding for thermostable aldehyde dehydrogenase from Bacillus
RT   stearothermophilus, and characterization of the enzyme.";
RL   J. Ferment. Bioeng. 76:161-167(1993).
CC   -!- FUNCTION: OXIDIZES SEVERAL ALIPHATIC ALDEHYDES, PARTICULARLY C6-
CC       ALIPHATIC ALDEHYDE AND HEXANAL, BUT DO NOT OXIDIZE BENZALDEHYDE.
CC       THE OPTIMUM TEMPERATURE IS AROUND 55 AND 60 DEGREES CELSIUS.
CC   -!- CATALYTIC ACTIVITY: An aldehyde + NAD(+) + H(2)O = an acid + NADH.
CC   -!- COFACTOR: REQUIRES EITHER NAD OR NADP AS A COENZYME.
CC   -!- PATHWAY: Ethanol utilization; second step.
CC   -!- SUBUNIT: HOMOTETRAMER.
CC   -!- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D13846; BAA02975.1; -.
DR   HSSP; P51977; 1BXS.
DR   InterPro; IPR002086; Aldehyde_dehydr.
DR   Pfam; PF00171; aldedh; 1.
DR   PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR   PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW   Oxidoreductase; NAD.
FT   NP_BIND     233    238       NAD (ADP PART) (BY SIMILARITY).
FT   ACT_SITE    255    255       BY SIMILARITY.
FT   ACT_SITE    289    289       BY SIMILARITY.
SQ   SEQUENCE   488 AA;  52915 MW;  29E824451985D9ED CRC64;

  Query Match             7.2%;  Score 88.5;  DB 1;  Length 488;
  Best Local Similarity   25.7%;  Pred. No. 3.8;
  Matches   46;  Conservative   23;  Mismatches   77;  Indels   33;  Gaps    8;

Qy         57 APAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERF--GVDRFLALLGSYGEGNV 114
              |||:      |||   |    || |  |         |   |  ||   :       | |:|
Db        168 APALVYGNTVVLKPASETAVTAAKVIEC-------FHEAGFPKGVVNMVC-----GSGSV 215

Qy        115 LVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTE-- 172
              :  |:     || :   | :  |:        | | :    || : | | :  |:| :
Db        216 VGQGIANHPDIDGVTFTGSNTVGKQVGRAAFERGAKY----QLEMGGKNPVIVAKDADLD 271

Qy        173 ---DALVSG--------CNGAAVALIERSLYEAHQ-RLDQSVRLLLHGGGVASLLPWLG 219
                  :  :|·       |   :   ||| :||  : :| : |: |  | |:     |:|
Db        272 LAVEGTISGGLRSTGQKCTATSRVFIEREVYEPFKAKLLERVKQLKIGNGL-DAETWMG 329


RESULT 12
CTPI_MYCLE
ID   CTPI_MYCLE     STANDARD;      PRT;  1609 AA.
AC   O53114;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Probable cation-transporting ATPase I (EC 3.6.3.-).
GN   CTPI OR ML2671 OR MLCB1913.02.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC       (E1-E2 ATPASES).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL022118; CAA17934.1; -.
DR   EMBL; AL583926; CAC32203.1; -.
DR   Leproma; ML2671; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR001454; Hlgnase/hydrlase.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00119; CATATPASE.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW   Complete proteome.
FT   TRANSMEM     30     50       POTENTIAL.
FT   TRANSMEM    176    196       POTENTIAL.
FT   TRANSMEM    238    258       POTENTIAL.
FT   TRANSMEM    357    377       POTENTIAL.
FT   TRANSMEM    641    661       POTENTIAL.
FT   TRANSMEM    673    693       POTENTIAL.
FT   TRANSMEM    778    798       POTENTIAL.
FT   TRANSMEM    921    941       POTENTIAL.
FT   TRANSMEM    969    989       POTENTIAL.
FT   TRANSMEM    997   1017       POTENTIAL.
FT   TRANSMEM   1396   1416       POTENTIAL.
FT   TRANSMEM   1426   1446       POTENTIAL.
FT   TRANSMEM   1542   1562       POTENTIAL.
FT   TRANSMEM   1573   1593       POTENTIAL.
FT   MOD_RES    1053   1053       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL      1335   1335       MAGNESIUM (BY SIMILARITY).
FT   METAL      1339   1339       MAGNESIUM (BY SIMILARITY).
```

```
SQ   SEQUENCE   1609 AA;  166871 MW;  58FA2079905E3995 CRC64;

  Query Match             7.2%;  Score 88;  DB 1;  Length 1609;
  Best Local Similarity   25.2%;  Pred. No. 16;
  Matches   68;  Conservative   29;  Mismatches   99;  Indels   74;  Gaps   12;

Qy         17 SLREGVIGPVTVLPYLTETMDAFA--------LQELPRGRVAYLASVAAPAITTHVLEVL 68
              :|   | | || || : :|: | |        :  |   |:  : :  | :|||     |
Db        191 ALSAAVTGWVTRLPAVPQTIRAAAALVNHQPRMVSLLESRLGRVGTDIALSITTAAASGL 250

Qy         69 KIHFEQVQVAATVAACAGVRI--AYAHPERFGVDR-------------FLALLGSYG--- 110
                 : |       || |:::  | || :|   ||             : :: | |
Db        251 T---QAVGTPLLDLACRGLQLSEAAAH-QRVWRDREPQLASPKRPQAPVVPVISSAGEKS 306

Qy        111 -------------EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQL 157
                          | : ||||      ||   | |   |   |   :    :| | |  |
Db        307 HAAGHNWTAAASNEASHLVVGGSIDAAID--TAKGSMKGPVESYVDSAANGSLIAAASAL 364

Qy        158 PLSGGNYLEFAEDTEDALVSGCNGAA-------VALIERSLYEAHQ---------RLDQS 201
                ||      ||   |:::|   ||        | : | |  | |          ||||
Db        365 LAGGG-----TEDAAGAILAGVPRAAHMGQQAFAATLGRGLANAGQLVLDPGALRRLDQV 419

Qy        202 VRLLLHGGGVASLLPWLGDVVHRPTLVLDG 231
                ::: |  :       ||  ||  |:  |
Db        420 KVVVIDGAALR------GD--HRAVLLARG 441


RESULT 13
GRK1_ECOLI
ID   GRK1_ECOLI     STANDARD;      PRT;   381 AA.
AC   P77364;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glycerate kinase 1 (EC 2.7.1.31).
GN   GLXK OR GLXB5 OR B0514.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RC   STRAIN=K12 / ECL1;
RX   MEDLINE=20069628; PubMed=10601204;
RA   Cusa E., Obradors N., Baldoma L., Badia J., Aguilar J.;
RT   "Genetic analysis of a chromosomal region containing genes required
RT   for assimilation of allantoin nitrogen and linked glyoxylate
RT   metabolism in Escherichia coli.";
RL   J. Bacteriol. 181:7479-7484(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + (R)-glycerate = ADP + 3-phospho-(R)-
CC       glycerate.
CC   -!- PATHWAY: GLYOXYLATE CATABOLISM.
CC   -!- INDUCTION: BY GLYOXYLATE.
CC   -!- SIMILARITY: BELONGS TO THE UPF0068 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000157; AAC73616.1; -.
DR   EMBL; U82664; AAB40266.1; -.
DR   EMBL; U89279; AAB93855.1; -.
DR   EcoGene; EG13621; glxK.
DR   InterPro; IPR004381; Cons_hypoth45.
DR   InterPro; IPR003747; Glycerate_kinase.
DR   Pfam; PF02595; DUF168; 1.
DR   TIGRFAMs; TIGR00045; Cons_hypoth45; 1.
KW   Transferase; Kinase; Complete proteome.
SQ   SEQUENCE   381 AA;  38734 MW;  0F902DF27DE63578 CRC64;

  Query Match             7.1%;  Score 87.5;  DB 1;  Length 381;
  Best Local Similarity   22.1%;  Pred. No. 3.4;
  Matches   54;  Conservative   25;  Mismatches   78;  Indels   87;  Gaps    9;

Qy         53 ASVAAPAITTHVLEVLKIHFEQVQV---------AATVAACAGVRIAYAHPERFGVDRFL 103
              | |||       |||     |:|             | :   |   :    ||:    |
Db         51 AMVAATGGNIVTLEVCGPMGEKVNAFYGLTGDGKTAVIEMAAASGLMLVAPEK----RNP 106

Qy        104 ALLGSYGEGNV-----------LVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHA 152
               |  |:| | :           :::|:| : |:|          | :  :  |  : | |
Db        107 LLASSFGTGELIRHALDNDIRHIILGIGGSATVD----------GGMGMAQALGVRFLDA 156

Qy        153 RAEQLPLSGGN------------------YLEFAEDTEDALVSGCNGAAVAL------I 187
                 : |  :|||                  ::| | | :: || |  |||
Db        157 DGQALAANGGNLARVASIEMDECDPRLANCHIEVACDVDNPLV-GARGAAAVFGPQKGAT 215

Qy        188 ERSLYEAHQRLDQSVRLL----------LHGGGVASLLPWLGDVVHRPTLVLDGLAIWAA 237
                 : |  | |    |:|          : ||| |                |: | ||
Db        216 PEMVEELEQGLQNYARVLQQQTEINVCQMAGGGAAG----------------GMGIAAA 258

Qy        238 VAAN 241
               |  |
Db        259 VFLN 262


RESULT 14
YPJA_ECOLI
ID   YPJA_ECOLI     STANDARD;      PRT;  1569 AA.
AC   P52143; P76610; P77017; P77019;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical outer membrane protein ypjA.
GN   YPJA OR B2647.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
```

```
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
CC   -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC   -!- SIMILARITY: STRONG, TO BORDETELLA PERTACTIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U36840; AAA79815.1; ALT_SEQ.
DR   EMBL; AE000350; AAC75695.1; -.
DR   EMBL; D90889; BAA16514.1; ALT_INIT.
DR   EMBL; D90890; BAA16518.1; ALT_INIT.
DR   EcoGene; EG13213; ypjA.
DR   InterPro; IPR004899; Pertact_sup.
DR   Pfam; PF03212; Pertactin; 1.
KW   Hypothetical protein; Outer membrane; Complete proteome.
SQ   SEQUENCE   1569 AA;  162774 MW;  0D006A9A32154596 CRC64;

  Query Match             7.0%;  Score 86.5;  DB 1;  Length 1569;
  Best Local Similarity   30.9%;  Pred. No. 21;
  Matches   50;  Conservative   19;  Mismatches   62;  Indels   31;  Gaps   10;

Qy         78 AATVAACAGVRIAYAHP-ERFGVDRFLA--LLGSYGEGNVLV----------VGVGTAL- 123
              |   :   |: |:  |    | :   ||  :|   | ||:||          || | |:
Db        917 ALQTSTANGIEISGTHQYGTFSISGNLATNMLLENG-GNLLVLAGTEARDSTVGKGGAMQ 975

Qy        124 TIDLLAANGCHLGGRISASPTLMR-----QALHARAEQLPLSGGNYLEFAEDTEDALVSG 178
               :   :|   : ||:     || |     ||| |||| | ::||  : :|    || |||
Db        976 NLGQDSATKVNSGGQY----TLGRSKDEFQAL-ARAEDLQVAGGTAIVYAGTLADASVSG 1030

Qy        179 CNGAAVALIERS-----LYEAHQRLDQSVRLLLHGGGVASLL 215
                |:   :  |        |   |:  |  | | | || : |
Db       1031 ATGSLSLMTPRDNVTPVKLEGAVRITDSATLTL-GNGVDTTL 1071


RESULT 15
MML7_MYCTU
ID   MML7_MYCTU     STANDARD;      PRT;   920 AA.
AC   P96289;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative membrane protein mmpL7.
GN   MMPL7 OR RV2942 OR MT3012 OR MTCY24G1.07C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MMPL FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z83858; CAB06107.1; -.
DR   EMBL; AE007123; AAK47339.1; -.
DR   TIGR; MT3012; -.
DR   TubercuList; Rv2942; -.
DR   InterPro; IPR004869; MMPL.
DR   Pfam; PF03176; MMPL; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM    210    230       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    271    291       POTENTIAL.
FT   TRANSMEM    311    331       POTENTIAL.
FT   TRANSMEM    344    364       POTENTIAL.
FT   TRANSMEM    389    409       POTENTIAL.
FT   TRANSMEM    761    781       POTENTIAL.
FT   TRANSMEM    790    810       POTENTIAL.
FT   TRANSMEM    822    842       POTENTIAL.
FT   TRANSMEM    864    884       POTENTIAL.
FT   TRANSMEM    888    908       POTENTIAL.
SQ   SEQUENCE   920 AA;  95121 MW;  019FCA9294A854D2 CRC64;

  Query Match             7.0%;  Score 85.5;  DB 1;  Length 920;
  Best Local Similarity   25.9%;  Pred. No. 14;
  Matches   62;  Conservative   22;  Mismatches   66;  Indels   89;  Gaps   13;

Qy         34 ETMDAF--ALQELPRGRVAYLASVAAPAITTH----------VLEVLKI 70
              |::||    |::||       ||:  ||||              || :|:
Db        175 ESLDAVRSVLRQLPPSE-GLRASIVVPAITNDMPMQITAWQSATIVTVAAVIAVLLLLRA 233

Qy         71 HFEQVQVAATVAACAGVRIAYAHP-------ERFGVDRFLALLGSYGEGNVLVVGVGTAL 123
                |: || |   | : :| | |         :| |       |:    || :|  ||
Db        234 RL-SVRAAAIVLLTADLSLAVAWPLAAVVRGHDWGTDSVF----SWTLAAVLTIGTITAA 288

Qy        124 TIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAA 183
              |: | |  |   |   ||:||  | :| | |  ||                       ||
Db        289 TM-LAARLGSDAGH--SAAPT-YRDSLPAFA--LP----------------------GAC 320

Qy        184 VALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAANV 242
              ||:    |  |           ||| | |                 || :: |:||::
Db        321 VAIFTGPLLLARTP-------ALHGVGTA-----------------GLGVFVALAASL 354


Search completed: June 24, 2003, 22:11:59
Job time : 5.86805 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:03:27 ; Search time 9.98363 Seconds
(without alignments)
2330.267 Million cell updates/sec

Title: US-09-813-453A-65
Perfect score: 1230
Sequence: 1 MNDWLFDLGNSRFKCASLRE.....HRPTLVLDGLAIWAAVAANV 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1230	100.0	242	2	A82637	conserved hypothet
2	200	16.3	592	2	B81009	BirA protein/Bvg a
3	198.5	16.1	592	2	H82031	probable biotin-[a
4	194.5	15.8	267	2	I40327	baf protein - Bord
5	193	15.7	248	2	H83111	hypothetical prote
6	140.5	11.4	261	2	B87489	transcription acti
7	139	11.3	265	2	T36391	hypothetical prote
8	130	10.6	273	2	E97293	probable transcrip
9	128.5	10.4	262	2	E75516	conserved hypothet
10	116	9.4	233	2	S66100	conserved hypothet
11	115.5	9.4	223	2	F64627	hypothetical prote
12	113.5	9.2	274	2	H86937	conserved hypothet
13	112	9.1	272	2	A70955	hypothetical prote
14	107.5	8.7	223	2	G71887	hypothetical prote
15	107.5	8.7	254	2	F83660	hypothetical prote
16	106	8.6	259	2	AF1102	conserved hypothet
17	104.5	8.5	276	2	AI2292	hypothetical prote
18	103	8.4	246	2	D72320	conserved hypothet
19	101	8.2	259	2	AF1464	conserved hypothet
20	99	8.0	257	2	S75559	hypothetical prote
21	95.5	7.8	464	2	C75474	conserved hypothet
22	95	7.7	229	2	E70465	hypothetical prote
23	94.5	7.7	655	2	JC2238	dnaK-type molecula
24	93.5	7.6	262	2	F70165	conserved hypothet
25	93.5	7.6	323	2	AH0903	octaprenyl-diphosp
26	93	7.6	399	2	G75610	imidazoloneropion
27	93	7.6	733	2	AB3636	sulfite reductase
28	92.5	7.5	397	2	I39579	nccB protein - Alc
29	92.5	7.5	631	2	S71508	dnaK-type molecula
30	91	7.4	407	2	D87023	hypothetical prote
31	90.5	7.4	260	2	F69278	conserved hypothet
32	90.5	7.4	400	2	D75331	hypothetical prote
33	90	7.3	301	2	T36884	hypothetical prote
34	89.5	7.3	209	2	H81382	hypothetical prote
35	89.5	7.3	328	2	AF2081	thioredoxin reduct
36	89.5	7.3	414	2	S60190	vicilin - Zamia fu
37	89.5	7.3	3491	2	T43231	probable 6-deoxyer
38	89	7.2	502	2	B70845	probable sygar tra
39	88.5	7.2	254	2	D85742	hypothetical prote
40	88.5	7.2	255	2	S04668	hypothetical prote
41	88.5	7.2	437	2	F70889	hypothetical prote
42	88.5	7.2	488	1	I39769	aldehyde dehydroge
43	88.5	7.2	859	2	B90768	probable tail leng
44	88.5	7.2	859	2	B85718	probable tail comp
45	88	7.2	273	2	D71326	conserved hypothet

ALIGNMENTS

RESULT 1
A82637
conserved hypothetical protein XF1795 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84603.1; GSPDB:G
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, F
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; F
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; L
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martin
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1795

Query Match 100.0%; Score 1230; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.5e-95;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60

Qy         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120

Qy        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180

Qy        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240

Qy        241 NV 242
              ||
Db        241 NV 242


RESULT 2
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain MC
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
  Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

  Query Match             16.3%;  Score 200;  DB 2;  Length 592;
  Best Local Similarity   30.6%;  Pred. No. 5.6e-09;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :|| ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |  | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571

Qy        231 GLAIW 235
              | |:
Db        572 NLVIY 576


RESULT 3
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidis (
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C;Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737911
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

  Query Match             16.1%;  Score 198.5;  DB 2;  Length 592;
  Best Local Similarity   30.7%;  Pred. No. 7.5e-09;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :|| ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |  | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578


RESULT 4
I40327
baf protein - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40327; S70669
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
J. Bacteriol. 177, 3801-3807, 1995
A;Title: Identification of a Bordetella pertussis regulatory factor required for tra
A;Reference number: I40327; MUID:95325323; PMID:7601846
A;Accession: I40327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229
R;Allen, A.; Maskell, D.
Mol. Microbiol. 19, 37-52, 1996
A;Title: The identification, cloning and mutagenesis of a genetic locus required for
A;Reference number: S70669; MUID:96419162; PMID:8821935
A;Accession: S70669
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 239-267 <ALL>
A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: baf

  Query Match             15.8%;  Score 194.5;  DB 2;  Length 267;
  Best Local Similarity   27.9%;  Pred. No. 6.3e-09;
  Matches   80;  Conservative   25;  Mismatches  105;  Indels   77;  Gaps   10;

Qy          5 LFDLGNSRFKCA--------SLREGVIGPVTVLPYLTETMDAFA----LQELPRG----- 47
              | | |||| |          : ||         |   : :|  |    |  |||
Db          4 LIDSGNSRLKVGWFDPDAPQAAREPA-------PVAFDNLDLDALGRWLATLPRRPQRAL 56

Qy         48 --RVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAA------------CAGVRIAYA 92
                 || ||   |                    :|||: |              |:|  |
Db         57 GVNVAGLARGEA------------------IAATLRAGGCDIRWLRAQPLAMGLRNGYR 97

Qy         93 HPERFGVDRFLALLGSYG-----EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMR 147
              :|:: | ||:  ::|           :||   ||| |:| :  :    || |   | :||
Db         98 NPDQLGADRWACMVGVLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMR 157

Qy        148 QALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLH 207
```

```
Db        158 GALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQAGAIVRQWLAGRQRYGQAPEIYVA 217

Qy        208 GGG-------VASLLPWLG---DVVHRPTL----VLDGLAIWAAVAA 240
Db        218 GGGWPEVRQEAERLLAVTGAAFGATPQPTYLDSPVLDGLAALAAQGA 264


RESULT 5
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

  Query Match             15.7%;  Score 193;  DB 2;  Length 248;
  Best Local Similarity   31.0%;  Pred. No. 7.7e-09;
  Matches   76;  Conservative   34;  Mismatches  111;  Indels   24;  Gaps    8;

Qy          7 DLGNSRFKCASLREGVIGPVTVLPYLTETMDAF-----ALQELPRGRVAYLASVAAPAIT 61
Db          6 DCGNSLIKWRVI-EGAAR--SVAGGLAESDDALVEQLTSQQALP-VRACRLVSVRSEQET 61

Qy         62 THVLEVLKIHFEQVQVAATVAA----CAGVRIAYAHPERFGVDRFLALLGSY--GEGNVL 115
Db         62 SQLVARLEQLF---PVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACL 118

Qy        116 VVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN------YLEFAE 169
Db        119 VIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQ 178

Qy        170 DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVL 229
b         179 ATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVF 238

Qy        230 DGLAI 234
Db        239 VGLAL 243


RESULT 6
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             11.4%;  Score 140.5;  DB 2;  Length 261;
  Best Local Similarity   27.2%;  Pred. No. 0.0002;
  Matches   49;  Conservative   30;  Mismatches   92;  Indels    9;  Gaps    4;

Qy         50 AYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS- 108
Db         60 AVIISSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAA 119

Qy        109 -YGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP-----LSGG 162
Db        120 MVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAG 179

Qy        163 NYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV 222
Db        180 NRI-VGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIA-TGGVASLFEGATDSI 237


RESULT 7
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE94.31c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             11.3%;  Score 139;  DB 2;  Length 265;
  Best Local Similarity   26.5%;  Pred. No. 0.00027;
  Matches   62;  Conservative   33;  Mismatches  107;  Indels   32;  Gaps    9;

Qy         16 ASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQV 75
Db         40 AVLLQGLMG----------MHPLLGDELGDGIDGIAICATVPSVLHELREVTRRYYGDV 88

Qy         76 -QVAATVAACAGVRIAYAHPERFGVDRFL---ALLGSYGEGNVLVVGVGTALTIDLLAAN 131
Db         89 PAVLVEPGVKTGVPILTDHPKEVGADRIINAVAAVELYG-GPAIVVDFGTATTFDAVSAR 147

Qy        132 GCHLGGRISASPTLMRQALHARAEQL-------PLS--GGNYLEFAEDTEDALVSGCNGA 182
Db        148 GEYIGGVIAPGIEISVEALGVKGAQLRKIEVARPRSVIGKNTV---EAMQSGIVYGFAGQ 204

Qy        183 AVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVV--HRPTLVLDGLAI 234
Db        205 VDGVVNRMARELADDPDDVT--VIATGGLAPMVLGESSVIDEHEPWLTLMGLRL 256


RESULT 8
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clost
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
```

```
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             10.6%;  Score 130;  DB 2;  Length 273;
  Best Local Similarity   24.5%;  Pred. No. 0.0016;
  Matches   53;  Conservative   43;  Mismatches  106;  Indels   14;  Gaps    7;

Qy         31 YLTETMDAFALQEL-PRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRI 89
              |  : |: |   :| |      : |   | |   :  ::: :|:   :        |: |
Db         50 YGIQVMNLFQQDKLDPTLVEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINI 109

Qy         90 AYAHPERFGVDRFLALLGSYG--EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMR 147
              | :|:  | || :  : ::   : :::::  ||| |   :  || :||| |     :
Db        110 KYDNPKEVGADRIVNAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSS 169

Qy        148 QALHARAEQLP---LSGGNYL---EFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQS 201
              :||  :| :||   |    |          :  :| |  |    ::|| : |   |   :
Db        170 EALFEKAAKLPRVELIKPAYAICKNTISSIQSGIVYGYIGQVRYIVER-MKEELQEEGEK 228

Qy        202 VRLLLHGGGVASLLPWLG---DVVHRPTLVLDGLAI 234
                |::  ||:| |:       ||:: | | |:|| |
Db        229 EPLVVATGGLAKLISEEAKNVDVIN-PFLTLEGLRI 263


RESULT 9
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: E75516
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <WHI>
A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g645814
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0461
A;Map position: 1
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             10.4%;  Score 128.5;  DB 2;  Length 262;
  Best Local Similarity   27.4%;  Pred. No. 0.002;
  Matches   58;  Conservative   25;  Mismatches  110;  Indels   19;  Gaps    6;

Qy         39 FALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFG 98
              | |   |  | | |:||| |    : |  || ||     | :      | :    |  |
Db         52 FTLAGAPIPRAAVLSSVAPPVGENYAL-ALKRHFMIDAFAVSAENLPDVTVELDTPGSVG 110

Qy         99 VDRFLALLGSY----GEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARA 154
              ||   | |:     |    :||  ||:   |::     ||| ::    :   || |||
Db        111 ADRLCNLFGAEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARA 170

Qy        155 EQLPLSGGNYLEFA--EDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGG--- 209
              :||       | |  ::|  || ||       :::  |      |      :  ||
Db        171 AKLPRITLQAPETAIGKNTVHALQSGLVFGYAEMVDGLLRRIRAELPGEAVAVATGGFSR 230

Qy        210 ---GVASLLPWLGDVVHRPTLVLDGLA-IWAA 237
                 |:   : :  :     || | ||  :||:
Db        231 TVQGICQEIDYYDE-----TLTLRGLVELWAS 257


RESULT 10
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S66100; E69740
R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis ch
A;Reference number: S65967; MUID:96051385; PMID:7584024
A;Accession: S66100
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <OGA>
A;Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g467
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 199
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e1182(
A;Experimental source: strain 168
C;Genetics:
A;Gene: yacB
A;Start codon: TTG
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             9.4%;  Score 116;  DB 2;  Length 233;
  Best Local Similarity   23.0%;  Pred. No. 0.019;
  Matches   47;  Conservative   34;  Mismatches   75;  Indels   48;  Gaps    7;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHV 64
              |||     |:     :|:|    | | :      |||: :          :  | : |
Db         46 LFDHSGLMFEQI---DGIIISSVVPPIM------FALERMCTKYFHIEPQIVGPGMKT-- 94

Qy         65 LEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRF---LALLGSYGEGNVLVVGVGT 121
                                  |: | | :|:  | ||    :| :  ||   ::||  ||
Db         95 --------------------GLNIKYDNPKEVGADRIVNAVAAIHLYG-NPLIVVDFGT 132

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLP---------LSGGNYLEFAEDTE 172
              | |   :  | ::|| |:   |:  :||::|| :||         : | | :     :
Db        133 ATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNTV---SAMQ 189

Qy        173 DALVSGCNGAAVALIERSLYEAHQ 196
                :: |  |    :::|  ::| |
Db        190 SGILFGYVGQVEGIVKRMKWQAKQ 213


RESULT 11
F64627
hypothetical protein HP0862 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
C;Accession: F64627
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKe
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey,
Nature 388, 539-547, 1997
```

A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64627
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <TOM>
A;Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07916.1; PID:g231400
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 9.4%; Score 115.5; DB 2; Length 223;
Best Local Similarity 31.5%; Pred. No. 0.02;
Matches 45; Conservative 22; Mismatches 55; Indels 21; Gaps 6;

```
Qy         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152
              |:|| :| |      | :||  |:|:||||:   | |||| |      : | :|     |
Db         85 GIDRQMACL---AVNNGVVVDAGSAITIDLI-KEGKHLGGCILPG---LAQYIHAYKKSA 137

Qy        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211
              :  : |    : ||   : | ||:  |   : :| |        | | :: :: | ||
Db        138 KILEQPFKALDSLEVLPKSTRDAVNYGMVLSVIACI--------QHLAKNQKIYLCGGDA 189

Qy        212 ASLLPWLGDVVHRPTLVLDGLAI 234
                |  :|   | :  || ||: |
Db        190 KYLSAFLPHSVCKERLVFDGMEI 212
```

RESULT 12
H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 9.2%; Score 113.5; DB 2; Length 274;
Best Local Similarity 26.5%; Pred. No. 0.038;
Matches 54; Conservative 30; Mismatches 95; Indels 25; Gaps 8;

```
Qy         52 LASVAAPAITTHVLEVLKIHFEQV-----QVAATVAACAGVRIAYAHPERFGVDRF---L 103
              ||  || :    ||  ::|  :|       |        |: :   :|:  | ||    |
Db         58 LAGAAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        104 ALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQL------ 157
              |    :|:   :||  |::: :|:::| |  ||| |:    :   |  ||:  |
Db        118 AAFHKFGQA-AIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELA 176

Qy        158 -PLS--GGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQ---RLDQSVRLLLHGGGV 211
               | |  | | :|     :  :| | |    |: |   :  :    |   | ::  |
Db        177 RPRSVVGKNTVEC---MQAGVVFGFAGLVDGLVGRMRQDVEEFSGDLGNRVAVVATGHTA 233

Qy        212 ASLLPWLGDVVHRPT-LVLDGLAI 234
                |||  |  |    | | || :
Db        234 PLLLPELHTVDHYDRHLTLHGLRL 257
```

RESULT 13
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70955
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70955
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-272 <COL>
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3600c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 9.1%; Score 112; DB 2; Length 272;
Best Local Similarity 25.3%; Pred. No. 0.051;
Matches 56; Conservative 32; Mismatches 107; Indels 26; Gaps 8;

```
Qy         32 LTETMDAFALQELPRGRVAYLASVAAPAITTHVLEVLKIHFEQV-----QVAATVAACAG 86
              |  |:|    ::  |     |   || :    ||  ::|  :|       |       |
Db         43 LALTIDGLIGEDSER-----LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTG 97

Qy         87 VRIAYAHPERFGVDRFLALLGSYGEGN--VLVVGVGTALTIDLLAANGCHLGGRISASPT 144
              : :   :|:  | || :  | :|       :||  |::: :|:::| |  ||| |:
Db         98 IPLLVDNPKEVGADRIVNCLAAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQ 157

Qy        145 LMRQALHARAEQL-------PLS--GGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAH 195
              :  | ||:  |       | |  | | :|     :   | |  |    |: |   :
Db        158 VSSDAAAARSAALRRVELARPRSVVGKNTVEC---MQAGAVFGFAGLVDGLVGRIREDVS 214

Qy        196 Q-RLDQSVRLLLHGGGVASLLPWLGDVVHRPT-LVLDGLAI 234
                :|  | ::  |    ||| |  | |    | | || :
Db        215 GFSVDHDVAIVATGHTAPLLLPELHTVDHYDQHLTLQGLRL 255
```

RESULT 14
G71887
hypothetical protein jhp0796 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Aug-2000
C;Accession: G71887
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: G71887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <ARN>
A;Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g41
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0796
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

Query Match 8.7%; Score 107.5; DB 2; Length 223;
Best Local Similarity 30.1%; Pred. No. 0.095;
Matches 43; Conservative 24; Mismatches 55; Indels 21; Gaps 6;

```
Qy         98 GVDRFLALLGSYGEGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMRQALH-----A 152
              |:|| :| |      | ::|  |:|:||||:   | |||| |      : | :|     |
Db         85 GIDRQMACLAVV---NGVIVDAGSAITIDLV-KEGKHLGGCILPG---LAQYVHAYKKSA 137

Qy        153 RAEQLPLSGGNYLE-FAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGV 211
              :  : |    : ||   ::| ||:  |   : :: |        | | :  :: | ||
```

```
Db        138 KILEQPFKALDSLEVLPKNTRDAVNYGMILSIISCI--------QHLAKDQKIYLCGGDA 189

Qy        212 ASLLPWLGDVVHRPTLVLDGLAI 234
                |  :|   | :  || ||: |
Db        190 KYLSAFLPHSVCKERLVFDGMEI 212


RESULT 15
F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83660
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A;Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03805.1; GSPDB:GN0
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0086
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             8.7%;  Score 107.5;  DB 2;  Length 254;
  Best Local Similarity   22.2%;  Pred. No. 0.11;
  Matches   49;  Conservative   41;  Mismatches   82;  Indels   49;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAITTHV 64
              |||    :|:     :|::    | | :      |:|::: :         :  | | |
Db         46 LFDHAGLQFQDI---DGIVISSVVPPMM------FSLEQMCKKYFHVTPMIIGPGIKT-- 94

Qy         65 LEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRF---LALLGSYGEGNVLVVGVGT 121
                              |: | | :|:  | ||    :| :  ||    :||  ||
Db         95 --------------------GLNIKYDNPKEVGADRIVNAVAAIELYGY-PAIVVDFGT 132

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAE-------DTEDA 174
               | |  |:     : || |:    :  :||: || :||      :| |:       :| |:
Db        133 ATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLP-----RIEIAKPKQVVGTNTIDS 187

Qy        175 LVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLL 215
              : ||     |: ::  :     : :   :::   ||:| |:
Db        188 MQSGIFYGYVSQVDGVVKRMKAQAESEPKVIA-TGGLAKLI 227


Search completed: June 24, 2003, 22:19:26
Db time : 11.9836 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:46:39 ; Search time 27.3931 Seconds
(without alignments)
1177.181 Million cell updates/sec

Title: US-09-813-453A-65
Perfect score: 1230
Sequence: 1 MNDWLFDLGNSRFKCASLRE.....HRPTLVLDGLAIWAAVAANV 242

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1230	100.0	242	23	AAU91180	Pantothenate kinas
2	200.5	16.3	389	21	AAY74909	Neisseria meningit
3	200	16.3	592	20	AAY38615	Neisseria meningit
4	200	16.3	592	21	AAY74912	Neisseria meningit
5	200	16.3	592	23	AAU91169	Pantothenate kinas
6	199.5	16.2	455	20	AAY38617	Neisseria gonorrho
7	199.5	16.2	455	21	AAY74908	Neisseria gonorrhe
8	199.5	16.2	460	23	AAU91167	Pantothenate kinas
9	199.5	16.2	592	20	AAY38618	Neisseria gonorrho
10	199.5	16.2	592	21	AAY74911	Neisseria gonorrhe
11	198.5	16.1	455	21	AAY74910	Neisseria meningit
12	198.5	16.1	592	20	AAY38616	Neisseria meningit
13	198.5	16.1	592	21	AAY74913	Neisseria meningit
14	198.5	16.1	592	23	AAU91166	Pantothenate kinas
15	194.5	15.8	267	23	AAU91162	Bordella pertussis
16	193	15.7	248	23	AAU91164	Pantothenate kinas
17	181.5	14.8	244	23	AAU91168	Pantothenate kinas
18	168.5	13.7	249	23	AAU91178	Pantothenate kinas
19	165.5	13.5	249	23	AAU91182	Pantothenate kinas
20	151.5	12.3	189	20	AAY38614	Neisseria meningit
21	144	11.7	257	23	AAU91174	Pantothenate kinas
22	143.5	11.7	219	23	AAU91176	Pantothenate kinas
23	143.5	11.7	241	23	AAU91179	Pantothenate kinas
24	140.5	11.4	260	23	AAU91173	Pantothenate kinas
25	139	11.3	265	23	AAU91151	Streptomyces coeli
26	134.5	10.9	258	22	AAU01243	B. subtilis novel
27	134.5	10.9	258	23	AAU91149	Bacillus subtilis
28	128.5	10.4	262	23	AAU91155	Deinococcus radiop
29	128	10.4	255	23	AAU91154	Geobacter sulfurre
30	125.5	10.2	258	23	AAU91153	Rhodobacter capsul
31	120	9.8	256	23	AAU91175	Pantothenate kinas
32	118	9.6	433	21	AAY80093	Permease fccI amin
33	117	9.5	262	23	AAU91170	Pantothenate kinas
34	116	9.4	233	23	AAU91163	Pantothenate kinas
35	115.5	9.4	223	19	AAW98422	H. pylori GHPO 344
36	115.5	9.4	223	23	AAU91161	Helicobacter pylor
37	115.5	9.4	223	23	AAU91181	Pantothenate kinas
38	115	9.3	258	23	AAU91172	Pantothenate kinas
39	112	9.1	272	22	AAG81225	Mycobacterium tube
40	112	9.1	272	23	AAU91152	Mycobacterium tube
41	108.5	8.8	250	23	AAU91150	Clostridium acetob
42	107.5	8.7	254	23	AAU91171	Pantothenate kinas
43	106	8.6	259	23	ABB47661	Listeria monocytog
44	105	8.5	498	22	AAU36184	Pseudomonas aerugi
45	104	8.5	212	23	AAU91177	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU91180
ID AAU91180 standard; Protein; 242 AA.
XX
AC AAU91180;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #18.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Xylessa fastidiosa.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54201.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein

```
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 111; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   242 AA;

  Query Match             100.0%;  Score 1230;  DB 23;  Length 242;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-124;
  Matches  242;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MNDWLFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYLASVAAPAI 60

Qy         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTHVLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGSYGEGNVLVVGVG 120

Qy        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCN 180

Qy        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVLDGLAIWAAVAA 240

Qy        241 NV 242
              ||
Db        241 NV 242


RESULT 2
AAY74909
ID   AAY74909 standard; Protein; 389 AA.
XX
AC   AAY74909;
XX
DT   21-MAR-2000  (first entry)
 X
 E   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53671.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 710; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   389 AA;

  Query Match             16.3%;  Score 200.5;  DB 21;  Length 389;
  Best Local Similarity   31.4%;  Pred. No. 6.1e-13;
  Matches   77;  Conservative   27;  Mismatches  118;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        139 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 198

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |   | :  :|   |  |:|  | |||  | ||:   |||  :    :||  ||
Db        199 VQEQLARKIEWLPSSA--QALFGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 256

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        257 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 316

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        317 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 369

Qy        231 GLAIW 235
              | |:
Db        370 NLVIY 374


RESULT 3
AAY38615
ID   AAY38615 standard; Protein; 592 AA.
XX
AC   AAY38615;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
```

```
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
T    Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             16.3%;  Score 200;  DB 20;  Length 592;
  Best Local Similarity   30.6%;  Pred. No. 1.3e-12;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | |  :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Y         122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571

Qy        231 GLAIW 235
              | |:
Db        572 NLVIY 576


RESULT 4
AAY74912
ID   AAY74912 standard; Protein; 592 AA.
XX
AC   AAY74912;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53674.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 714; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             16.3%;  Score 200;  DB 21;  Length 592;
  Best Local Similarity   30.6%;  Pred. No. 1.3e-12;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | |  :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571

Qy        231 GLAIW 235
              | |:
Db        572 NLVIY 576
```

```
RESULT 5
AAU91169
ID   AAU91169 standard; Protein; 592 AA.
XX
AC   AAU91169;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #7.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria meningitidis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54190.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 6; Page 96-98; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   592 AA;

  Query Match           16.3%;  Score 200;  DB 23;  Length 592;
  Best Local Similarity   30.6%;  Pred. No. 1.3e-12;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
                | |      :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |     |    | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571

Qy        231 GLAIW 235
               | |:
Db        572 NLVIY 576


RESULT 6
AAY38617
ID   AAY38617 standard; Protein; 455 AA.
XX
AC   AAY38617;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   455 AA;

  Query Match           16.2%;  Score 199.5;  DB 20;  Length 455;
  Best Local Similarity   30.3%;  Pred. No. 9.8e-13;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | : |||| | | :  |    |   ||   :       |   | | :  :|   :
Db        205 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 264

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
                | |      :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        265 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 321

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |     |    | :|: ||
Db        322 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 381

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        382 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 437
```

```
Qy        229 LDGL 232
              : ||
Db        438 IHGL 441


RESULT 7
AAY74908
ID   AAY74908 standard; Protein; 455 AA.
XX
AC   AAY74908;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1290.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53670.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 709; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match           16.2%;  Score 199.5;  DB 21;  Length 455;
  Best Local Similarity   30.3%;  Pred. No. 9.8e-13;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | : |||| | | :  |    |   ||   :       |   | |  :  :|   :
Db        205 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 264

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        265 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 321

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        322 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 381

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        382 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 437

Qy        229 LDGL 232
              : ||
Db        438 IHGL 441


RESULT 8
AAU91167
ID   AAU91167 standard; Protein; 460 AA.
XX
AC   AAU91167;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #5.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54188.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 8; Page 92-94; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   460 AA;

  Query Match           16.2%;  Score 199.5;  DB 23;  Length 460;
  Best Local Similarity   30.3%;  Pred. No. 1e-12;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;
```

```
Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | : |||| | | :  |    |   ||   :        |  | | :  :|  :
Db        210 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 269

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        270 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 326

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |   | :|: ||
Db        327 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 386

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :      | ::: |||   ||  ||  :| : ·|    ||
Db        387 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 442

Qy        229 LDGL 232
              : ||
Db        443 IHGL 446
```

RESULT 9
AAY38618
ID AAY38618 standard; Protein; 592 AA.
XX
AC AAY38618;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigenic protein encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI; 1999-327407/27.
DR N-PSDB; AAZ12075.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 175; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 592 AA;

```
  Query Match            16.2%;  Score 199.5;  DB 20;  Length 592;
  Best Local Similarity  30.3%;  Pred. No. 1.4e-12;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | : |||| | | :  |    |   ||   :        |  | | :  :|  :
Db        342 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :      | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578
```

RESULT 10
AAY74911
ID AAY74911 standard; Protein; 592 AA.
XX
AC AAY74911;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1296.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53673.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 713; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

```
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             16.2%;  Score 199.5;  DB 21;  Length 592;
  Best Local Similarity   30.3%;  Pred. No. 1.4e-12;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | : |||| | | :  |    |   ||   :       |   | |  :  :|   :
Db        342 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |   |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578


RESULT 11
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
X
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             16.1%;  Score 198.5;  DB 21;  Length 455;
  Best Local Similarity   30.7%;  Pred. No. 1.3e-12;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | |  :  :|
Db        205 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 264

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        265 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 321

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |   |   | :|: ||
Db        322 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 381

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        382 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 437

Qy        229 LDGL 232
              : ||
Db        438 IHGL 441


RESULT 12
AAY38616
ID   AAY38616 standard; Protein; 592 AA.
XX
AC   AAY38616;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
```

```
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             16.1%;  Score 198.5;  DB 20;  Length 592;
  Best Local Similarity   30.7%;  Pred. No. 1.8e-12;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578


RESULT 13
AAY74913
ID   AAY74913 standard; Protein; 592 AA.
XX
AC   AAY74913;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1300.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53675.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 715; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             16.1%;  Score 198.5;  DB 21;  Length 592;
  Best Local Similarity   30.7%;  Pred. No. 1.8e-12;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | | :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578


RESULT 14
```

AAU91166
ID AAU91166 standard; Protein; 592 AA.
XX
AC AAU91166;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #4.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria meningitidis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 85-86; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 592 AA;

Query Match 16.1%; Score 198.5; DB 23; Length 592;
Best Local Similarity 30.7%; Pred. No. 1.8e-12;
Matches 75; Conservative 29; Mismatches 117; Indels 23; Gaps 8;

```
Qy          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63
              | | |||| | | :  |    |   ||   :       |   | |  :  :|
Db        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401

Qy         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121
              | |  |    :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458

Qy        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181
              |:|:|  |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518

Qy        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574

Qy        229 LDGL 232
              : ||
Db        575 IHGL 578
```

RESULT 15
AAU91162
ID AAU91162 standard; Protein; 267 AA.
XX
AC AAU91162;
XX
DT 05-JUN-2002 (first entry)
XX
DE Bordella pertussis pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Bordetella pertussis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 79-80; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 267 AA;

Query Match 15.8%; Score 194.5; DB 23; Length 267;
Best Local Similarity 27.9%; Pred. No. 1.6e-12;
Matches 80; Conservative 25; Mismatches 105; Indels 77; Gaps 10;

```
Qy          5 LFDLGNSRFKCA--------SLREGVIGPVTVLPYLTETMDAFA----LQELPRG----- 47
              | | |||| |           : ||          |   : :|  |    |  |||
Db          4 LIDSGNSRLKVGWFDPDAPQAAREPA-------PVAFDNLDLDALGRWLATLPRRPQRAL 56

Qy         48 --RVAYLASVAAPAITTHVLEVLKIHFEQVQVAATVAA-------------CAGVRIAYA 92
                || ||   |                    :|||: |               |:|  |
Db         57 GVNVAGLARGEA------------------IAATLRAGGCDIRWLRAQPLAMGLRNGYR 97

Qy         93 HPERFGVDRFLALLGSYG-----EGNVLVVGVGTALTIDLLAANGCHLGGRISASPTLMR 147
              :|:: | ||:  ::|           :||   ||| |:| :  :    || |   | :||
Db         98 NPDQLGADRWACMVGVLARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMR 157

Qy        148 QALHARAEQLPLSGGNYLEFAEDTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLH 207
               ||      |||: |   ::  ||  |: ||   |    | |      ||  |:  : :
Db        158 GALAYGTAHLPLADGLVADYPIDTHQAIASGIAAAQAGAIVRQWLAGRQRYGQAPEIYVA 217

Qy        208 GGG-------VASLLPWLG---DVVHRPTL----VLDGLAIWAAVAA 240
              |||          ||   |       :||     ||||||  ||  |
Db        218 GGGWPEVRQEAERLLAVTGAAFGATPQPTYLDSPVLDGLAALAAQGA 264
```

Search completed: June 24, 2003, 22:08:42
Job time : 28.3931 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 11.7629 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1335	100.0	258	9	US-09-813-453A-2	Sequence 2, Appli
2	1114	83.4	233	9	US-09-813-453A-17	Sequence 17, Appl
3	1046	78.4	258	9	US-09-813-453A-49	Sequence 49, Appl
4	1034	77.5	262	9	US-09-813-453A-45	Sequence 45, Appl
5	987	73.9	254	9	US-09-813-453A-47	Sequence 47, Appl
6	794.5	59.5	256	9	US-09-813-453A-55	Sequence 55, Appl
7	756	56.6	255	9	US-09-813-453A-7	Sequence 7, Appli
8	655	49.1	265	9	US-09-813-453A-4	Sequence 4, Appli
9	646.5	48.4	250	9	US-09-813-453A-3	Sequence 3, Appli
10	577	43.2	260	9	US-09-813-453A-51	Sequence 51, Appl
11	537	40.2	258	9	US-09-813-453A-6	Sequence 6, Appli
12	515	38.6	219	9	US-09-813-453A-57	Sequence 57, Appl
13	494.5	37.0	272	9	US-09-813-453A-5	Sequence 5, Appli
14	493.5	37.0	272	9	US-09-712-363-276	Sequence 276, App
15	428	32.1	262	9	US-09-813-453A-8	Sequence 8, Appli
16	405.5	30.4	246	9	US-09-813-453A-9	Sequence 9, Appli
17	338.5	25.4	212	9	US-09-813-453A-59	Sequence 59, Appl
18	327.5	24.5	273	9	US-09-813-453A-10	Sequence 10, Appl
19	319	23.9	257	9	US-09-813-453A-53	Sequence 53, Appl
20	271.5	20.3	262	9	US-09-813-453A-11	Sequence 11, Appl
21	207	15.5	244	9	US-09-813-453A-41	Sequence 41, Appl
22	203	15.2	241	9	US-09-813-453A-63	Sequence 63, Appl
23	163	12.2	249	9	US-09-813-453A-70	Sequence 70, Appl
24	163	12.2	257	9	US-09-813-453A-13	Sequence 13, Appl
25	154.5	11.6	229	9	US-09-813-453A-12	Sequence 12, Appl
26	154	11.5	249	9	US-09-813-453A-61	Sequence 61, Appl
27	151	11.3	460	9	US-09-813-453A-39	Sequence 39, Appl
28	150	11.2	592	9	US-09-813-453A-22	Sequence 22, Appl
29	150	11.2	592	9	US-09-813-453A-43	Sequence 43, Appl
30	134.5	10.1	242	9	US-09-813-453A-65	Sequence 65, Appl
31	133	10.0	248	9	US-09-813-453A-20	Sequence 20, Appl
32	129.5	9.7	267	9	US-09-813-453A-15	Sequence 15, Appl
33	109.5	8.2	223	9	US-09-895-913A-74	Sequence 74, Appl
34	109.5	8.2	223	9	US-09-813-453A-14	Sequence 14, Appl
35	109.5	8.2	223	9	US-09-813-453A-67	Sequence 67, Appl
36	109	8.2	209	9	US-09-813-453A-21	Sequence 21, Appl
37	80	6.0	449	10	US-09-815-242-5474	Sequence 5474, Ap
38	80	6.0	449	10	US-09-815-242-12348	Sequence 12348, A
39	80	6.0	449	10	US-09-815-242-12793	Sequence 12793, A
40	79	5.9	636	10	US-09-205-658-160	Sequence 160, App
41	78.5	5.9	337	9	US-10-075-846-12	Sequence 12, Appl
42	78.5	5.9	776	10	US-09-815-242-13811	Sequence 13811, A
43	78.5	5.9	1057	10	US-09-815-242-5798	Sequence 5798, Ap
44	78.5	5.9	1107	10	US-09-815-242-12815	Sequence 12815, A
45	78.5	5.9	1198	10	US-09-815-242-12446	Sequence 12446, A

ALIGNMENTS

RESULT 1
US-09-813-453A-2
; Sequence 2, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-813-453A-2

Query Match 100.0%; Score 1335; DB 9; Length 258;
Best Local Similarity 100.0%; Pred. No. 1.5e-129;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240

Qy        241 LTLKGLELIYERNRVGSV 258
              ||||||||||||||||||
Db        241 LTLKGLELIYERNRVGSV 258


RESULT 2
US-09-813-453A-17
; Sequence 17, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
;   LENGTH: 233
;   TYPE: PRT
;   ORGANISM: Bacillus subtilis
US-09-813-453A-17

  Query Match            83.4%;  Score 1114;  DB 9;  Length 233;
  Best Local Similarity  99.5%;  Pred. No. 7.9e-107;
  Matches  213;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214
              |||||||||||||||||||||||||||||||||
Db        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214


RESULT 3
US-09-813-453A-49
; Sequence 49, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49
;   LENGTH: 258
;   TYPE: PRT
;   ORGANISM: Bacillus stearothermophilus
US-09-813-453A-49

  Query Match            78.4%;  Score 1046;  DB 9;  Length 258;
  Best Local Similarity  78.3%;  Pred. No. 9.3e-100;
  Matches  198;  Conservative   32;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :: |:|||||||||||||   :|::|||||||| |||||:||::::| :| || | | ||
Db          1 MIFVLDVGNTNTVLGVYDGDELKHHWRIETSRSKTEDEYGMMIKALLNHVGLQFSDIRGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIKPLIVGPGIKTGLDIKYDNPREVGADRIVNAVAGIHL 120
              |||||||||||||||| |||||:| |||||:||||:||||||:|||||||||||| |||
Db         61 IISSVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREVGADRIVNAVAGIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||:|||:||||||||||||:|:|||||||||||| ||||||::|||||||| |||:||
Db        121 YGSPLIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRIEIARPDDII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              |||||||||:|||:|||||||||| ||| ::|   |||||||||||||:||| ||:||||
Db        181 GKNTVSAMQAGILYGYVGQVEGIVSRMKAKSKIPPKVIATGGLAPLIASESDIIDVVDPF 240

Qy        241 LTLKGLELIYERN 253
              ||| ||:|:||:|
Db        241 LTLTGLKLLYEKN 253


RESULT 4
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
US-09-813-453A-45

  Query Match            77.5%;  Score 1034;  DB 9;  Length 262;
  Best Local Similarity  75.1%;  Pred. No. 1.6e-98;
  Matches  190;  Conservative   34;  Mismatches   29;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :: |:|||||| |||| :|:|  |||:|| |||||||:||::: | :| || || : ||
Db          1 MIFVLDVGNTNAVLGVFEEGELRQHWRMETDRHKTEDEYGMLVKQLLEHEGLSFEDVKGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |:|||||||||||||| ||| |:| :||||:|||||||:||:|||||||||||| |||
Db         61 IVSSVVPPIMFALERMCEKYFKIKPLVVGPGIKTGLNIKYENPREVGADRIVNAVAGIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||:|||:||||||||||||:| | |||| | ||| || |||||||||||||||:| :::
Db        121 YGSPLIIVDFGTATTYCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRIEITKPSSVV 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||||||||||||:|||||||||||||| :|||: |||||||| ||: ||: ||:||||
Db        181 GKNTVSAMQSGILYGYVGQVEGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIDVVDPF 240

Qy        241 LTLKGLELIYERN 253
```

```
                  |||||| ::|||||
Db        241 LTLKGLYMLYERN 253


RESULT 5
US-09-813-453A-47
; Sequence 47, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 47
;   LENGTH: 254
;   TYPE: PRT
;   ORGANISM: Bacillus halodurans
US-09-813-453A-47

  Query Match             73.9%;  Score 987;  DB 9;  Length 254;
  Best Local Similarity   72.3%;  Pred. No. 1.1e-93;
  Matches  183;  Conservative   33;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||||||||||||||| |  | :|||: ||| |||||: | :||||||:|| |: ||||
Db          1 MILVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              :||||||:||:||:|| ||||: | |:|||:||||||||||||||||||||||||||| |
Db         61 VISSVVPPMMFSLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRIVNAVAAIEL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || | |||||||||||:| | |  :|:|| ||||| ||:|||:  |:||||:|: :|   |
Db        121 YGYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQVV 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              | ||: :||||| :||| ||:|:||||| ||: : ||||||||| ||  ||: ||::| |
Db        181 GTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240

Qy        241 LTLKGLELIYERN 253
              ||||||:|||::|
Db        241 LTLKGLQLIYKKN 253


RESULT 6
US-09-813-453A-55
; Sequence 55, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 55
;   LENGTH: 256
;   TYPE: PRT
;   ORGANISM: Clostridium difficile
US-09-813-453A-55

  Query Match             59.5%;  Score 794.5;  DB 9;  Length 256;
  Best Local Similarity   60.8%;  Pred. No. 7.9e-74;
  Matches  155;  Conservative   40;  Mismatches   59;  Indels    1;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :||| |||||| |||:|   ||  :|||:| | || ||:|::: :|||:  :    || :
Db          1 MLLVFDVGNTNMVLGIYKGDKLVNYWRIKTDREKTSDEYGILISNLFDYDNVNISDIDDV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| :| :||  | ||   :| |||||:||||||||||||:|||||||||| |
Db         61 IISSVVPNVMHSLENFCIKYCKKQPLIVGPGIKTGLNIKYDNPKQVGADRIVNAVAGIEK 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || | |:|||||||:| | |  :|:|| ||||| ||:|||:  |:||||:|: :|   |
Db        121 YGAPSILVDFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPRVELAKPGMTI 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAK-QDLKVIATGGLAPLIANESDCIDIVDP 239
              |:|||||||||::|||| |: |:  || :    |:|||||||| |||:|:  || ||
Db        181 CKSTVSAMQSGIIYGYVGLVDKIISIMKKELNCDDVKVIATGGLAKLIASETKSIDYVDG 240

Qy        240 FLTLKGLELIYERNR 254
              ||||:|| :|||:|:
Db        241 FLTLEGLRIIYEKNQ 255


RESULT 7
US-09-813-453A-7
; Sequence 7, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Geobacter sulfurreducens
US-09-813-453A-7

  Query Match             56.6%;  Score 756;  DB 9;  Length 255;
  Best Local Similarity   57.5%;  Pred. No. 7.3e-70;
  Matches  146;  Conservative   46;  Mismatches   62;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|||||||||| |||:|   :|   ||: | : :| ||:|:::  ||  :||  :||  :
Db          1 MLLVIDVGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGILINELFRLAGLGLDQIRAV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||:   |||:   || : | :||||:|||: |:||||:|||||||||||
Db         61 IISSVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPIQYDNPREVGADRIVNAVAGYEK 120

Qy        121 YGNPLIVVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPRVDIIRPSAII 180
              |   ||:|||||||: |::   :| |||||||: |||||: ||:||||::| ||  ||
Db        121 YRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPRVDIIRPSAII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              :|||::||:|| :|||| |: || ||| ::|   :|||||||| ||| ||  |: |: :
Db        181 ARNTVNSMQAGIYYGYVGLVDEIVTRMKAESKDAPRVIATGGLASLIAPESKTIEAVEEY 240
```

```
Qy        241 LTLKGLELIYERNR 254
              |||:|| ::|||||
Db        241 LTLEGLRILYERNR 254


RESULT 8
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

  Query Match             49.1%;  Score 655;  DB 9;  Length 265;
  Best Local Similarity   51.0%;  Pred. No. 2e-59;
  Matches  133;  Conservative   46;  Mismatches   74;  Indels    8;  Gaps    3;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ---- 56
              :|| ||||||:||||::    :  |||| |   :| ||  ::|: |     |: ::
Db          1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDG 60

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              |||| | : || ::  |  :  :|:   | : | ||:|||: |  |:|||||||||:|||
Db         61 IDGIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPILTDHPKEVGADRIINAV 120

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||: ||| | ||||||||||:  :    :|:|| ||||| || |||  : |:| :||: |
Db        121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVAR 180

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQA---KQDLKVIATGGLAPLIANESD 232
              | ::|||||| ||||||::|: |||:|:| ||  :      |: |||||||||::  ||
Db        181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGGLAPMVLGESS 240

Qy        233 CIDIVDPFLTLKGLELIYERN 253
              ||  :|:||| || |:||||
Db        241 VIDEHEPWLTLMGLRLVYERN 261


RESULT 9
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

  Query Match             48.4%;  Score 646.5;  DB 9;  Length 250;
  Best Local Similarity   53.9%;  Pred. No. 1.4e-58;
  Matches  125;  Conservative   46;  Mismatches   56;  Indels    5;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::||:|||||| |||:|:| ||   ||: |  :: ||:|: : :||   |   ::|:
Db         18 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGV 77

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| ||::|| |  ||| | | :||||:|||:|||||||||||||||||||||  :
Db         78 IISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVAAHEI 137

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              |  ||::|||||||:| : ||  |:|||| ||| :|:|||: :||||||:|: :|   |
Db        138 YKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKPAYAI 197

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLK-----VIATGGLAPLI 227
              |||:|::|||||:: |: ||: : :::|       :     |:|||||| ||
Db        198 CKNTISSIQSGIVYRYLRQVKYLFEKLKENLPDGRRTRTSLVLATGGLAKLI 249


RESULT 10
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-813-453A-51

  Query Match             43.2%;  Score 577;  DB 9;  Length 260;
  Best Local Similarity   44.4%;  Pred. No. 2.1e-51;
  Matches  114;  Conservative   50;  Mismatches   91;  Indels    2;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|| |: |||||:  ::        || |   :| ||: : |  |    || |  || :
Db          1 MLLAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||  :| |  :  :||::|| ::|   | |:::: : | | ||||:|||: |  :
Db         61 IISSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRP--DN 178
              |  ||:|:| |||||:  :  :  : || ||||| :| :||:  |||||||  | ||  :
Db        121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180

Qy        179 IIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVD 238
              |:| :|||||||: :||:  :||:| |:| :  : : ||||||:| |    :| ||  |
Db        181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFD 240

Qy        239 PFLTLKGLELIYERNRV 255
              ||::||  || || :
Db        241 SDLTIRGLLEIYRRNTI 257
```

RESULT 11
US-09-813-453A-6
; Sequence 6, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Rhodobacter capsulatus
US-09-813-453A-6

Query Match 40.2%; Score 537; DB 9; Length 258;
Best Local Similarity 43.9%; Pred. No. 2.8e-47;
Matches 112; Conservative 46; Mismatches 95; Indels 2; Gaps 2;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|| || |||||| |:        ||| |   :| ||: : | :|    ||   :|
Db          1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLNTLMQLKGLQ-GRISEA 59

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVG-PGMKTGLNIKYDNPKEVGADRIVNAVAAIH 119
              ||||  | ::| |  :| :||   | :|| || : :  : |    || ||:|| ||
Db         60 IISSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119

Qy        120 LYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 179
               :|  ||||||||||:  :  :  |:|| |||| :| |||:  || || :::|:|  :
Db        120 RHGGDLIVVDFGTATTFDVVAPDGAYIGGVIAPGVNLSLEALHMAAAALPHVDVTKPQGV 179

Qy        180 IGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDP 239
              || |||: :|||: :||:| |||||:::: :  : :|||||||| |      |  | |:
Db        180 IGTNTVACIQSGVYWGYIGLVEGIVRQIRMERDRPMKVIATGGLASLFDLGFDLFDKVED 239

Qy        240 FLTLKGLELIYERNR 254
               ||: || ||:: |:
Db        240 DLTMHGLRLIFDYNK 254
```

RESULT 12
US-09-813-453A-57
; Sequence 57, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 57
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Dehalococcoides ethenogenes

US-09-813-453A-57

Query Match 38.6%; Score 515; DB 9; Length 219;
Best Local Similarity 45.8%; Pred. No. 4.1e-45;
Matches 97; Conservative 45; Mismatches 70; Indels 0; Gaps 0;

```
Qy          2 LLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGII 61
              |: :|:|||:  :|::   ||  :| : :   :  ||:  :|  |  |:|:  |::: :|
Db          5 LVAVDIGNTSVNIGIFEGEKLLANWHLGSVAQRMADEYASLLLGLLQHAGIHPEELNRVI 64

Qy         62 ISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121
              : |||||:    | :   ||   | :|| |:|:|: :: |||:||||||||| ||  ||
Db         65 MCSVVPPLTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAAAARVLY 124

Qy        122 GNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIG 181
                 |:|| |||||:  : |   |:||||||||  | :|:  : :|||:||| ||  :||
Db        125 PGACIIVDMGTATTFDTLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAKVIG 184

Qy        182 KNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQ 213
               |||||||||| |||:| || :|:|:: :  |
Db        185 SNTVSAMQSGIYFGYIGLVEELVRRIQTELGQ 216
```

RESULT 13
US-09-813-453A-5
; Sequence 5, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-813-453A-5

Query Match 37.0%; Score 494.5; DB 9; Length 272;
Best Local Similarity 38.8%; Pred. No. 7.3e-43;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

```
Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  : :  |        |:
Db          1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : .: ||::||: :  ||||||||||||| :
Db         58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||   :    ||||||::      :    :::|||||||: :|::|  :|:| | |:|: |
Db        118 AAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
              | :::|||||  ||:| :||: | |:|:| |::     :    |: ::|||  |||:  |
Db        178 PRSVVGKNTVECMQAGAVFGFAGLVDGLVGRIREDVSGFSVDHDVAIVATGHTAPLLLPE 237

Qy        231 SDCIDIVDPFLTLKGLELIYERN 253
               :|  |  |||:|| |::|||
Db        238 LHTVDHYDQHLTLQGLRLVFERN 260
```

RESULT 14

```
US-09-712-363-276
; Sequence 276, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
;  APPLICANT: Eisenberg, David
;  APPLICANT:  Rotstein, Sergio H.
;  APPLICANT:  Marcotte, Edward M.
;  TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
;  TITLE OF INVENTION:  INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
;  FILE REFERENCE: 07419-032001
;  CURRENT APPLICATION NUMBER: US/09/712,363
;  CURRENT FILING DATE:  2000-11-13
;  PRIOR APPLICATION NUMBER: PCT/US00/02246
;  PRIOR FILING DATE: 2000-01-28
;  PRIOR APPLICATION NUMBER: 60/179,531
;  PRIOR FILING DATE: 2000-02-01
;  PRIOR APPLICATION NUMBER: 60/117,844
;  PRIOR FILING DATE: 1999-01-29
;  PRIOR APPLICATION NUMBER: 60/118,206,
;  PRIOR FILING DATE: 1999-02-01
;  PRIOR APPLICATION NUMBER: 60/126,593
;  PRIOR FILING DATE: 1999-03-26
;  PRIOR APPLICATION NUMBER: 60/134,093
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/134,092
;  PRIOR FILING DATE: 1999-05-14
;  PRIOR APPLICATION NUMBER: 60/165,124
;  PRIOR FILING DATE: 1999-11-12
;  PRIOR APPLICATION NUMBER: 60/165,086
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 292
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 276
;   LENGTH: 272
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
US-09-712-363-276

  Query Match             37.0%;  Score 493.5;  DB 9;  Length 272;
  Best Local Similarity   38.8%;  Pred. No. 9.2e-43;
  Matches  102;  Conservative   55;  Mismatches   93;  Indels   13;  Gaps    4;

Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  : :  |        |:
Db          1 VLLAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : : ||::||: :  ||||||||||||| :
Db         58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

 y        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||   :    ||||||::     :    :::|||||||: :|::|  :|:| | |:|: |
Db        118 AAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
              | :::|||||  ||:| :||: | |:|:| |::     :    |: ::|||  |||:  |
Db        178 PRSVVGKNTVECMQAGAVFGFAGLVDGLVGRIREDVSGFSVDHDVAIVATGHTAPLLLPE 237

Qy        231 SDCIDIVDPFLTLKGLELIYERN 253
                :|  |  |||:|| |::|||
Db        238 LHTVDHYDQHLTLQGLRLVFERN 260


RESULT 15
US-09-813-453A-8
; Sequence 8, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Deinococcus radiopugnans
US-09-813-453A-8

  Query Match             32.1%;  Score 428;  DB 9;  Length 262;
  Best Local Similarity   37.4%;  Pred. No. 5e-36;
  Matches   98;  Conservative   51;  Mismatches   99;  Indels   14;  Gaps    6;

Qy          2 LLVIDVGNTNTVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              || :|:||| ||||:    | | : ||| |:|    |:  : |  ||  :|    :
Db          6 LLAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAGAPIPR--AA 63

Qy         61 IISSVVPPI----MFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVA 116
              ::||| ||:       ||:|    :| |:   |        : :: | |  |||||: |
Db         64 VLSSVAPPVGENYALALKR----HFMIDAFAVSAENLPDVTVELDTPGSVGADRLCNLFG 119

Qy        117 AIHLYG--NPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEIT 174
              |    |  :  :||||||:| :  :   ::::|| :| |  :| :||::|||||||| :
Db        120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy        175 RPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCI 234
              |:  ||||||  |:|||::|||   |:|:::|:: :   :   :|||| :  :      |
Db        180 APETAIGKNTVHALQSGLVFGYAEMVDGLLRRIRAELPGEAVAVATGGFSRTVQGICQEI 239

Qy        235 DIVDPFLTLKGL-ELIYERNRV 255
              |  |  |||:|| ||   |: |
Db        240 DYYDETLTLRGLVELWASRSEV 261


Search completed: June 24, 2003, 22:23:49
Job time : 12.7629 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 7.53698 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	83.5	6.3	207	3	US-08-549-515-9	Sequence 9, Appli
2	79	5.9	342	4	US-08-096-181A-12	Sequence 12, Appl
3	79	5.9	342	4	US-08-096-181A-14	Sequence 14, Appl
4	79	5.9	342	5	PCT-US94-08326-12	Sequence 12, Appl
5	79	5.9	342	5	PCT-US94-08326-14	Sequence 14, Appl
6	79	5.9	361	4	US-08-096-181A-8	Sequence 8, Appli
7	79	5.9	361	5	PCT-US94-08326-8	Sequence 8, Appli
8	79	5.9	363	4	US-08-096-181A-10	Sequence 10, Appl
9	79	5.9	363	5	PCT-US94-08326-10	Sequence 10, Appl
10	79	5.9	923	3	US-08-936-135-6	Sequence 6, Appli
11	78	5.8	1012	1	US-07-944-943-2	Sequence 2, Appli
12	78	5.8	1012	1	US-08-219-262B-2	Sequence 2, Appli
13	78	5.8	1012	3	US-09-031-655-2	Sequence 2, Appli
14	78	5.8	1213	2	US-08-937-102-2	Sequence 2, Appli
15	78	5.8	3898	2	US-08-876-991-2	Sequence 2, Appli
16	78	5.8	3898	2	US-09-059-853-2	Sequence 2, Appli
17	78	5.8	3898	4	US-08-750-717-2	Sequence 2, Appli
18	77	5.8	627	4	US-09-071-035-452	Sequence 452, App
19	77	5.8	1313	4	US-09-071-035-450	Sequence 450, App
20	77	5.8	1313	4	US-09-071-035-454	Sequence 454, App
21	76	5.7	922	4	US-09-116-473-4	Sequence 4, Appli
22	75.5	5.7	443	4	US-09-594-193-9	Sequence 9, Appli
23	75.5	5.7	1215	4	US-09-134-001C-5319	Sequence 5319, Ap
24	75	5.6	553	6	5310678-1	Patent No. 5310678
25	75	5.6	1012	1	US-08-219-262B-9	Sequence 9, Appli
26	75	5.6	1012	3	US-09-031-655-9	Sequence 9, Appli
27	74.5	5.6	457	4	US-09-134-001C-3838	Sequence 3838, Ap
28	74	5.5	413	4	US-09-147-928-2	Sequence 2, Appli
29	73	5.5	724	4	US-08-914-999-4	Sequence 4, Appli
30	73	5.5	2319	1	US-08-212-133A-8	Sequence 8, Appli
31	73	5.5	2319	1	US-08-474-503-6	Sequence 6, Appli
32	73	5.5	2319	2	US-08-670-707A-6	Sequence 6, Appli
33	73	5.5	2319	4	US-09-037-601-6	Sequence 6, Appli
34	73	5.5	2319	4	US-09-315-179-6	Sequence 6, Appli
35	73	5.5	2319	4	US-09-523-656-28	Sequence 28, Appl
36	73	5.5	2319	5	PCT-US94-13200-6	Sequence 6, Appli
37	72.5	5.4	396	2	US-08-878-989-16	Sequence 16, Appl
38	72.5	5.4	396	4	US-09-272-796-16	Sequence 16, Appl
39	72.5	5.4	396	4	US-09-344-700-2	Sequence 2, Appli
40	72.5	5.4	820	1	US-08-291-896-2	Sequence 2, Appli
41	72.5	5.4	820	2	US-08-485-278-2	Sequence 2, Appli
42	72.5	5.4	1148	4	US-09-106-075A-90	Sequence 90, Appl
43	72.5	5.4	1612	1	US-08-169-927-2	Sequence 2, Appli
44	72	5.4	314	4	US-09-134-001C-2983	Sequence 2983, Ap
45	72	5.4	361	3	US-09-032-372-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-549-515-9
; Sequence 9, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Haemophilus Influenzae
; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 Unviersity Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-549-515-9

Query Match 6.3%; Score 83.5; DB 3; Length 207;
Best Local Similarity 24.4%; Pred. No. 0.11;
Matches 47; Conservative 27; Mismatches 66; Indels 53; Gaps 11;

```
Qy         15 GVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIIISSVVPPIMFALE 74
              ||:|    |: : | : ||       :    || | :: || ::          |
Db         52 GVWHQNVFAYY--LSISCNHCEDP----ACTKVCPSGAMHKREDGFVVVD---------E 96

Qy         75 RMC--TKYFHIEPQIVGPGMKTGLNIKYDNPK------EVGADRIVNAVAAIHLYGNPLI 126
```

```
Db         97 DVCIGCRYCHMACPYGAP--------QYNETKGHMTKCDGCYDRVAEGKKPICVESCPLR 148

Qy        127 VVDFGTATTYCYIDENKQYMG--GAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNT 184

Db        149 ALDFGP------IDELRKKHGDLAAVAP-----------RA--LPRAHFTKP-NIVIKPN 188

Qy        185 VSAMQSGILFGYV 197

Db        189 ANSRPTGDTTGYL 201


RESULT 2
US-08-096-181A-12
; Sequence 12, Application US/08096181A
; Patent No. 6153406
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,181A
;      FILING DATE:  23-Jul-1993
;      CLASSIFICATION:  424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-096-181A-12

  Query Match             5.9%;  Score 79;  DB 4;  Length 342;
  Best Local Similarity   20.1%;  Pred. No. 0.83;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73

Db        100 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 156

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133

Db        157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 203

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP-------------R 170

Db        204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 263

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195

Db        264 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 297


RESULT 3
US-08-096-181A-14
; Sequence 14, Application US/08096181A
; Patent No. 6153406
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,181A
;      FILING DATE:  23-Jul-1993
;      CLASSIFICATION:  424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-096-181A-14

  Query Match             5.9%;  Score 79;  DB 4;  Length 342;
  Best Local Similarity   20.1%;  Pred. No. 0.83;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73

Db        100 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 156

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133

Db        157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 203

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP-------------R 170

Db        204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 263

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195

Db        264 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 297


RESULT 4
PCT-US94-08326-12
; Sequence 12, Application PC/TUS9408326
;  GENERAL INFORMATION:
;    APPLICANT:  North American Vaccine, Inc.
;    APPLICANT:  12103 Indian Creek Court
;    APPLICANT:  Beltsville, MD 20705
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level
;    TITLE OF INVENTION:  Expression,
```

```
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane
;    TITLE OF INVENTION:  Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08326
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,181
;      FILING DATE:  23-JULY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REFERENCE/DOCKET NUMBER:  1438.001PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-08326-12

  Query Match             5.9%;  Score 79;  DB 5;  Length 342;
  Best Local Similarity   20.1%;  Pred. No. 0.83;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73
              ::  | ||         : |:|::  | :     :  |  |: |||::: :     : |
Db        100 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 156

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133
              :|   |  :  |      :: |   : :|  :|||    | : |   ||
Db        157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 203

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP------------R 170
              | | |   ||:|| : | :|         |: :| :: |::                 :
Db        204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 263

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
                |:    |: |     : :|       :  :|||
Db        264 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 297


RESULT 5
PCT-US94-08326-14
; Sequence 14, Application PC/TUS9408326
;  GENERAL INFORMATION:
;    APPLICANT:  North American Vaccine, Inc.
;    APPLICANT:  12103 Indian Creek Court
;    APPLICANT:  Beltsville, MD 20705
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level
;    TITLE OF INVENTION:  Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane
;    TITLE OF INVENTION:  Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08326
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,181
;      FILING DATE:  23-JULY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REFERENCE/DOCKET NUMBER:  1438.001PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  342 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-08326-14

  Query Match             5.9%;  Score 79;  DB 5;  Length 342;
  Best Local Similarity   20.1%;  Pred. No. 0.83;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73
              ::  | ||         : |:|::  | :     :  |  |: |||::: :     : |
Db        100 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 156

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133
              :|   |  :  |      :: |   : :|  :|||    | : |   ||
Db        157 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 203

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP------------R 170
              | | |   ||:|| : | :|         |: :| :: |::                 :
Db        204 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 263

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
                |:    |: |     : :|       :  :|||
Db        264 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 297


RESULT 6
US-08-096-181A-8
; Sequence 8, Application US/08096181A
; Patent No. 6153406
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
```

```
;       COUNTRY:  U.S.A.
;       ZIP:  20005-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/096,181A
;       FILING DATE:  23-Jul-1993
;       CLASSIFICATION:  424
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  361 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-096-181A-8

  Query Match             5.9%;  Score 79;  DB 4;  Length 361;
  Best Local Similarity   20.1%;  Pred. No. 0.91;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73
              ::  | ||         : |:|::  | :      :  |  |: |||::: :      : |
Db        119 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 175

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133
              :|   |  :  |       :: |   : :|  :|||    | : |   ||
Db        176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 222

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP------------R 170
              | | |   ||:|| : | :|         |: :| :: |::              :
Db        223 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 282

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
               |:    |: |     : :|       :  :|||
Db        283 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 316


RESULT 7
PCT-US94-08326-8
; Sequence 8, Application PC/TUS9408326
;  GENERAL INFORMATION:
;    APPLICANT:  North American Vaccine, Inc.
;    APPLICANT:  12103 Indian Creek Court
;    APPLICANT:  Beltsville, MD 20705
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level
;    TITLE OF INVENTION:  Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane
;    TITLE OF INVENTION:  Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08326
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,181
;      FILING DATE:  23-JULY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REFERENCE/DOCKET NUMBER:  1438.001PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  361 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-08326-8

  Query Match             5.9%;  Score 79;  DB 5;  Length 361;
  Best Local Similarity   20.1%;  Pred. No. 0.91;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73
              ::  | ||         : |:|::  | :      :  |  |: |||::: :      : |
Db        119 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 175

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133
              :|   |  :  |       :: |   : :|  :|||    | : |   ||
Db        176 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 222

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP------------R 170
              | | |   ||:|| : | :|         |: :| :: |::              :
Db        223 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 282

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
               |:    |: |     : :|       :  :|||
Db        283 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 316


RESULT 8
US-08-096-181A-10
; Sequence 10, Application US/08096181A
; Patent No. 6153406
;  GENERAL INFORMATION:
;    APPLICANT:  Tai, Joseph Y.
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,181A
;      FILING DATE:  23-Jul-1993
;      CLASSIFICATION:  424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  10:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-096-181A-10

  Query Match             5.9%;  Score 79;  DB 4;  Length 363;
  Best Local Similarity   20.1%;  Pred. No. 0.91;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73
              ::  | ||         : |:|::  | :      :  |  |: |||::: :      : |
Db        121 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 177

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133
              :|   |  :  |      :: |   : :|  :|||    | : |   ||
Db        178 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 224

 y        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP-------------R 170
              | | |   ||:|| : | :|         |: :| :: |::                  :
Db        225 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 284

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
               |:    |: |     : :|       :  :|||
Db        285 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 318


RESULT 9
PCT-US94-08326-10
; Sequence 10, Application PC/TUS9408326
;  GENERAL INFORMATION:
;    APPLICANT:  North American Vaccine, Inc.
;    APPLICANT:  12103 Indian Creek Court
;    APPLICANT:  Beltsville, MD 20705
;    APPLICANT:  Pullen, Jeffrey K.
;    APPLICANT:  Soper, Thomas S.
;    APPLICANT:  Liang, Shu-Mei
;    TITLE OF INVENTION:  A Method For The High Level
;    TITLE OF INVENTION:  Expression,
;    TITLE OF INVENTION:  Purification And Refolding Of The Outer Membrane
;    TITLE OF INVENTION:  Protein
;    TITLE OF INVENTION:  P2 From Haemophilus Influenzae Type b
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;      STREET:  1100 New York Avenue, Suite 600
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3934
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08326
;      FILING DATE:  Herewith
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/096,181
;      FILING DATE:  23-JULY-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Esmond, Robert W.
;      REFERENCE/DOCKET NUMBER:  1438.001PC01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 371-2600
;      TELEFAX:  (202) 371-2540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  363 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US94-08326-10

  Query Match             5.9%;  Score 79;  DB 5;  Length 363;
  Best Local Similarity   20.1%;  Pred. No. 0.91;
  Matches   43;  Conservative   38;  Mismatches   71;  Indels   62;  Gaps   10;

Qy         28 IETSRHKT---------EDEFGMILRSLF-----DHSGLMFEQIDGIIISSVVPPIMFAL 73
              ::  | ||         : |:|::  | :      :  |  |: |||::: :      : |
Db        121 VKLGRAKTIADGITSAEDKEYGVLNNSDYIPTSGNTVGYTFKGIDGLVLGA---NYLLAQ 177

Qy         74 ERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTA 133
              :|   |  :  |      :: |   : :|  :|||    | : |   ||
Db        178 KREGAKGENKRPNDKAGEVRIG---EINNGIQVGAKYDANDIVAKIAYGR---------- 224

Qy        134 TTYCY--IDENKQYMGGAIAP--------GITISTEALYSRAAKLP-------------R 170
              | | |   ||:|| : | :|         |: :| :: |::                  :
Db        225 TNYKYNESDEHKQQLNGVLATLGYRFSDLGLLVSLDSGYAKTKNYKIKHEKRYFVSPGFQ 284

Qy        171 IEITRPDNIIG-----KNTV----SAMQSGILFG 195
               |:    |: |     : :|       :  :|||
Db        285 YELMEDTNVYGNFKYERTSVDQGEKTREQAVLFG 318


RESULT 10
US-08-936-135-6
; Sequence 6, Application US/08936135
; Patent No. 6054293
;  GENERAL INFORMATION:
;    APPLICANT:  Tessier-Lavigne, Marc
;    APPLICANT:  He, Zhigang
;    APPLICANT:  Chen, Hang
;    TITLE OF INVENTION:  Semaphorin Receptors
;    NUMBER OF SEQUENCES:  26
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/936,135
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  UC97-288-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  923 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-936-135-6

  Query Match             5.9%;  Score 79;  DB 3;  Length 923;
  Best Local Similarity   21.4%;  Pred. No. 4;
  Matches   47;  Conservative   32;  Mismatches   81;  Indels   60;  Gaps   11;

Qy          8 GNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSG---LMFEQID----GI 60
              |:  |     ||   :  | : ||:            : ||:|    :: | |    |
```

```
Db        655 GSHKTFCHWEHDSHAQLRWSVLTSK----------TGPIQDHTGDGNFIYSQADENQKGK 704

Qy         61 IISSVVPPIMFALERMC-TKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVA--- 116
              :   | | :       | | ::|:    ||      : ::|  |:|   |::|  |
Db        705 VARLVSPVVYSQSSAHCMTFWYHMSGSHVG---TLRVKLRYQKPEEY--DQLVWMVVGHQ 759

Qy        117 ------------AIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYS 163
                          :: ||     |:  |        :  |  :||     |:|:
Db        760 GDHWKEGRVLLHKSLKLYQ----VIFEG--------EIGKGNLGGIAVDDISINNHISQE 807

Qy        164 RAAKLPRIEITRPDNIIGKNT-VSAMQSGILFGYVGQVEG 202
                ||       | ::  ||| :   ::|   || |: ||
Db        808 DCAK--------PTDLDKKNTEIKIDETGSTPGYEGEGEG 839


RESULT 11
US-07-944-943-2
; Sequence 2, Application US/07944943
; Patent No. 5518724
;  GENERAL INFORMATION:
     APPLICANT:  SNYDER, DAVID B.
     APPLICANT:  VAKHARIA, VIKRAM
;    TITLE OF INVENTION:  NOVEL INFECTIOUS BURSAL DISEASE VIRUS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, McCLELLAND, MAIER & NEUSTADT,
;      ADDRESSEE:  P.C.
;      STREET:  1755 S. Jefferson Davis Highway, Suite 400
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/944,943
;      FILING DATE:  19920915
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kelber, Steven B.
;      REGISTRATION NUMBER:  30,073
;      REFERENCE/DOCKET NUMBER:  2284-029-0 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
   INFORMATION FOR SEQ ID NO:  2:
,    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1012 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-944-943-2

  Query Match             5.8%;  Score 78;  DB 1;  Length 1012;
  Best Local Similarity   23.8%;  Pred. No. 6.1;
  Matches   39;  Conservative   22;  Mismatches   67;  Indels   36;  Gaps    6;

Qy         97 NIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDF-GTATTYCYIDENKQYMGGAIAPGIT 155
              ||       || : :       : |  : :: | |||     :  |          |:|
Db        233 NIDAITSLSVGGELVFKTSVQSLVLGATIYLIGFDGTAVITRAVAANN---------GLT 283

Qy        156 ISTEALYSRAAKLPRIEITRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDL 215
                |: |      :|  |||:|       :::::  |:    | :||   :| | |   |
Db        284 AGTDNLMPFNLVIPTNEITQP--------ITSIKLKIVTSKSGGLEG--DQMSWSASGSL 333

Qy        216 KVIATGGLAPLIANESDCIDIVDPFLTLKGLELI-YERNRVGSV 258
                |   ||  |               |: : |: |||   |||
Db        334 AVTIHGGNYP--------------GALRPVTLVAYERVATGSV 362


RESULT 12
US-08-219-262B-2
; Sequence 2, Application US/08219262B
; Patent No. 5788970
;  GENERAL INFORMATION:
;    APPLICANT:  VAKHARIA, VIKRAM
;    APPLICANT:  SNYDER, DAVID B
;    APPLICANT:  MENGEL-WHERSAT, STEPHANIE A
;    TITLE OF INVENTION:  CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
;    TITLE OF INVENTION:  cDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
;    TITLE OF INVENTION:  THEREON
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/219,262B
;      FILING DATE:  29-MAR-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2747-047-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1012 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Infectious bursal disease virus
;      STRAIN:  DS326
US-08-219-262B-2

  Query Match             5.8%;  Score 78;  DB 1;  Length 1012;
  Best Local Similarity   23.8%;  Pred. No. 6.1;
  Matches   39;  Conservative   22;  Mismatches   67;  Indels   36;  Gaps    6;

Qy         97 NIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDF-GTATTYCYIDENKQYMGGAIAPGIT 155
              ||       || : :       : |  : :: | |||     :  |          |:|
Db        233 NIDAITSLSVGGELVFKTSVQSLVLGATIYLIGFDGTAVITRAVAANN---------GLT 283

Qy        156 ISTEALYSRAAKLPRIEITRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDL 215
                |: |      :|  |||:|       :::::  |:    | :||   :| | |   |
Db        284 AGTDNLMPFNLVIPTNEITQP--------ITSIKLKIVTSKSGGLEG--DQMSWSASGSL 333

Qy        216 KVIATGGLAPLIANESDCIDIVDPFLTLKGLELI-YERNRVGSV 258
                |   ||  |               |: : |: |||   |||
Db        334 AVTIHGGNYP--------------GALRPVTLVAYERVATGSV 362


RESULT 13
US-09-031-655-2
; Sequence 2, Application US/09031655
; Patent No. 6017759
;  GENERAL INFORMATION:
;    APPLICANT:  VAKHARIA, VIKRAM
;    APPLICANT:  SNYDER, DAVID B
```

```
;    APPLICANT:  MENGEL-WHERSAT, STEPHANIE A
;    TITLE OF INVENTION:  CHIMERIC INFECTIOUS BURSAL DISEASE VIRUS
;    TITLE OF INVENTION:  cDNA CLONES, EXPRESSION PRODUCTS AND VACCINES BASED
;    TITLE OF INVENTION:  THEREON
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
;      STREET:  1755 S. JEFFERSON DAVID HIGHWAY, FOURTH FLOOR
;      CITY:  ARLINGTON
;      STATE:  VIRGINIA
;      COUNTRY:  USA
;      ZIP:  22202
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/031,655
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/219,262
;      FILING DATE:  29-MAR-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OBLON, NORMAN F
;      REGISTRATION NUMBER:  24,618
;      REFERENCE/DOCKET NUMBER:  2747-047-27
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 413-3000
;      TELEFAX:  (703) 413-2220
;      TELEX:  248855 OPAT UR
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1012 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Infectious bursal disease virus
;      STRAIN:  DS326
US-09-031-655-2

  Query Match             5.8%;  Score 78;  DB 3;  Length 1012;
  Best Local Similarity   23.8%;  Pred. No. 6.1;
  Matches   39;  Conservative   22;  Mismatches   67;  Indels   36;  Gaps   6;

Qy         97 NIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDF-GTATTYCYIDENKQYMGGAIAPGIT 155
              ||       || : :        : |  : :: | |||     :  |          |:|
Db        233 NIDAITSLSVGGELVFKTSVQSLVLGATIYLIGFDGTAVITRAVAANN---------GLT 283

Qy        156 ISTEALYSRAAKLPRIEITRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDL 215
               |: |      :|  |||:|        :::::  |:    | :||   :| | |   |
Db        284 AGTDNLMPFNLVIPTNEITQP--------ITSIKLKIVTSKSGGLEG--DQMSWSASGSL 333

Qy        216 KVIATGGLAPLIANESDCIDIVDPFLTLKGLELI-YERNRVGSV 258
               |   ||  |                    |: : |: |||   |||
Db        334 AVTIHGGNYP--------------GALRPVTLVAYERVATGSV 362


RESULT 14
US-08-937-102-2
; Sequence 2, Application US/08937102A
; Patent No. 5965134
; GENERAL INFORMATION:
;  APPLICANT: Thiel, Heinz-Jurgen
;  APPLICANT: Elpers, Knut
;  APPLICANT: Pauly, Thomas
;  TITLE OF INVENTION: T Cell Stimulating Protein of Pestivirus
;  FILE REFERENCE: I/94108
;  CURRENT APPLICATION NUMBER: US/08/937,102A
;  CURRENT FILING DATE: 1997-09-24
;  EARLIER APPLICATION NUMBER: 08/693,247
;  EARLIER FILING DATE: 1996-08-16
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1213
;   TYPE: PRT
;   ORGANISM: Classical Swine Fever Virus
US-08-937-102-2

  Query Match             5.8%;  Score 78;  DB 2;  Length 1213;
  Best Local Similarity   20.7%;  Pred. No. 8.1;
  Matches   45;  Conservative   35;  Mismatches   73;  Indels   64;  Gaps   11;

Qy         20 GKLEYHW-RIETSRHKTEDEFGMILR--------SLFDHSGLMFEQID---GIIISSVVP 67
              |  :||:  ::  |:  ||   :           ||::   ||  |::    ::||  :|
Db        983 GSKDYHYDLLQAQRYGIEDGINITKSFREMNYDWSLYEEDSLMITQLEILNNLLISEELP 1042

Qy         68 PIMFALERMCTKYFHIEP-----------QIVGPGMKTG-LNIKYDNPKEVGADRIVNA 114
                 |:: :  :  | ||           :: | :: | :   |||   : | :: :
Db       1043 ---MAVKNIMARTDHPEPIQLAYNSYETQVPVLFPKIRNGEVTDTYDNYTFLNARKLGDD 1099

Qy        115 VAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIA-----PGITISTEALYSRAAKLP 169
              |                    | |  |::      :      ||   : ||   || |
Db       1100 V-----------------PPYVYATEDEDLAVELLGLDWPDPGNQGTVEA--GRALK-- 1137

Qy        170 RIEITRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKR 206
                    ::| :|       |||||  : : ||
Db       1138 --------QVVGLSTAENALLVALFGYVG-YQALSKR 1165


RESULT 15
US-08-876-991-2
; Sequence 2, Application US/08876991
; Patent No. 5925360
; GENERAL INFORMATION:
;    APPLICANT:  Gregor Meyers, Tillmann R menapf,
;    APPLICANT:  Heinz-J rgen Thiel
;    TITLE OF INVENTION:  Hog cholera virus vaccine and diagnostic
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Organon Teknika Corporation
;      ADDRESSEE:  Biotechnology Research Institute
;      STREET:  1330-A Piccard Drive
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  U.S.A.
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/876,991
;      FILING DATE:  16-JUN-1997
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/747,577
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/650,584
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/469,702
;      FILING DATE:
;      APPLICATION NUMBER:  US/08/123,596
;      FILING DATE:
;      APPLICATION NUMBER:  07/797,554
;      FILING DATE:  22-NOV-1991
;      APPLICATION NUMBER:  US 07/494,991
;      FILING DATE:  16-MAR-1990
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  William M. Blackstone
```

```
;      REGISTRATION NUMBER:  29,772
;      REFERENCE/DOCKET NUMBER:
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 258-5200
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3898 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-876-991-2

  Query Match             5.8%;  Score 78;  DB 2;  Length 3898;
  Best Local Similarity   20.7%;  Pred. No. 51;
  Matches   45;  Conservative   35;  Mismatches   73;  Indels   64;  Gaps   11;

Qy         20 GKLEYHW-RIETSRHKTEDEFGMILR--------SLFDHSGLMFEQID---GIIISSVVP 67
              |  :||:  ::  |:  ||   :          ||::   ||  |::    ::||  :|
Db       2113 GSKDYHYDLLQAQRYGIEDGINITKSFREMNYDWSLYEEDSLMITQLEILNNLLISEELP 2172

Qy         68 PIMFALERMCTKYFHIEP------------QIVGPGMKTG-LNIKYDNPKEVGADRIVNA 114
                |:: :  :  | ||             :: | :: | :   |||   : | :: :
Db       2173 ---MAVKNIMARTDHPEPIQLAYNSYETQVPVLFPKIRNGEVTDTYDNYTFLNARKLGDD 2229

Qy        115 VAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIA-----PGITISTEALYSRAAKLP 169
              |                    | |  |::      :      ||   : ||   || |
Db       2230 V-----------------PPYVYATEDEDLAVELLGLDWPDPGNQGTVEA--GRALK-- 2267

Qy        170 RIEITRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKR 206
                  ::| :|       ||||||  : : ||
Db       2268 --------QVVGLSTAENALLVALFGYVG-YQALSKR 2295


Search completed: June 24, 2003, 21:49:12
Job time : 9.53698 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 21.0426 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1036	77.6	258	2	Q9F985	Q9f985 bacillus st
2	987	73.9	254	16	Q9KGH5	Q9kgh5 bacillus ha
3	852	63.8	259	16	Q8YAC5	Q8yac5 listeria mo
4	847	63.4	259	16	Q92F54	Q92f54 listeria in
5	775	58.1	273	16	Q97EB4	Q97eb4 clostridium
6	773	57.9	255	16	Q8R7M2	Q8r7m2 thermoanaer
7	737	55.2	259	16	Q8XHL5	Q8xhl5 clostridium
8	655	49.1	265	16	Q9X8N6	Q9x8n6 streptomyce
9	577	43.2	261	16	Q9A6Z1	Q9a6z1 caulobacter
10	519.5	38.9	274	16	Q9CD56	Q9cd56 mycobacteri
11	494.5	37.0	272	16	O06282	O06282 mycobacteri
12	458	34.3	256	16	Q8RFE4	Q8rfe4 fusobacteri
13	428	32.1	262	16	Q9RX54	Q9rx54 deinococcus
14	405.5	30.4	246	16	Q9WZY5	Q9wzy5 thermotoga
15	338.5	25.4	212	2	O32514	O32514 desulfovibr
16	327.5	24.5	273	16	O83446	O83446 treponema p
17	271.5	20.3	262	16	O51477	O51477 borrelia bu
18	171	12.8	295	16	Q8Y2M4	Q8y2m4 ralstonia s
19	170	12.7	276	16	Q8YQD7	Q8yqd7 anabaena sp
20	166	12.4	56	2	P94305	P94305 bacillus ps
21	163	12.2	257	16	P74045	P74045 synechocyst
22	154.5	11.6	229	16	O67753	O67753 aquifex aeo
23	150	11.2	592	16	Q9JXF1	Q9jxf1 neisseria m
24	150	11.2	592	16	Q9JWI7	Q9jwi7 neisseria m
25	134.5	10.1	242	16	Q9PCI4	Q9pci4 xylella fas
26	133	10.0	248	16	Q9HWC1	Q9hwc1 pseudomonas
27	117	8.8	224	16	Q98Q93	Q98q93 mycoplasma
28	116.5	8.7	223	16	Q9ZKY6	Q9zky6 helicobacte
29	109.5	8.2	223	16	O25533	O25533 helicobacte
30	109	8.2	209	16	Q9PIA9	Q9pia9 campylobact
31	104.5	7.8	597	17	O30225	O30225 archaeoglob
32	100	7.5	318	10	Q9LVY3	Q9lvy3 arabidopsis
33	95	7.1	520	16	Q8U6P6	Q8u6p6 agrobacteri
34	94.5	7.1	858	12	Q9QTD8	Q9qtd8 marek's dis
35	93.5	7.0	895	17	Q972N0	Q972n0 sulfolobus
36	93	7.0	1014	5	O96276	O96276 plasmodium
37	91.5	6.9	485	16	Q8RFP6	Q8rfp6 fusobacteri
38	91.5	6.9	828	16	Q9PBB9	Q9pbb9 xylella fas
39	90	6.7	676	12	Q91MV8	Q91mv8 lumpy skin
40	89.5	6.7	357	16	Q8R832	Q8r832 thermoanaer
41	89.5	6.7	858	12	Q9E1I5	Q9e1i5 meleagrid h
42	89.5	6.7	858	12	Q9DPS9	Q9dps9 meleagrid h
43	87.5	6.6	205	16	Q8ZPJ8	Q8zpj8 salmonella
44	87.5	6.6	287	16	Q98QY2	Q98qy2 mycoplasma
45	87	6.5	336	16	Q8Z7S9	Q8z7s9 salmonella

ALIGNMENTS

RESULT 1
Q9F985
ID Q9F985 PRELIMINARY; PRT; 258 AA.
AC Q9F985;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative 32 kDa replication protein.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Geobacillus.
OX NCBI_TaxID=1422;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V;
RA Vasquez C., Pichuantes S., Saavedra C.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF198621; AAG28531.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00671; baf; 1.
SQ SEQUENCE 258 AA; 28101 MW; 507B55D695095855 CRC64;

Query Match 77.6%; Score 1036; DB 2; Length 258;
Best Local Similarity 77.9%; Pred. No. 3.9e-84;
Matches 197; Conservative 31; Mismatches 25; Indels 0; Gaps 0;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :: |:||||||||||||   :|::||||||| |||||:|| :::| :| || |  ||||
Db          1 MIFVLDVGNTNTVLGVYDGDELKHHWRIETSRGKTEDEYGMTIKALLNHVGLQFSDIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||||||||||||| |||||:| |||||:||||||||||:||||||||||| |||
Db         61 IISSVVPPIMFALERMCLKYFHIKPIIVGPGIKTGLNIKYDNPREVGADRIVNAVAGIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||:|||:||||||||||||:|:||||||||||| ||||||::|||||||| |||:||
Db        121 YGSPLIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRIEIARPDDII 180
```

```
Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              |||||||||:|||:|||||||||| ||| ::    |||||||| |||:||: |||||||
Db        181 GKNTVSAMQAGILYGYVGQVEGIVSRMKAKSPVPPKVIATGGLASLIASESNVIDIVDPF 240

Qy        241 LTLKGLELIYERN 253
              ||| ||:::||:|
Db        241 LTLTGLKILYEKN 253


RESULT 2
Q9KGH5
ID   Q9KGH5      PRELIMINARY;      PRT;   254 AA.
AC   Q9KGH5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein BH0086.
GN   BH0086.
OS   Bacillus halodurans.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Bacillaceae; Bacillus.
OX   NCBI_TaxID=86665;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C-125 / JCM 9153;
RX   MEDLINE=20512582; PubMed=11058132;
RA   Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA   Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA   Horikoshi K.;
RT   "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT   halodurans and genomic sequence comparison with Bacillus subtilis.";
RL   Nucleic Acids Res. 28:4317-4331(2000).
DR   EMBL; AP001507; BAB03805.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   254 AA;  27907 MW;  0E3B557BA7DAC176 CRC64;

  Query Match           73.9%;  Score 987;  DB 16;  Length 254;
  Best Local Similarity  72.3%;  Pred. No. 8.6e-80;
  Matches  183;  Conservative   33;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||||||||||||| |  | :|||: ||| |||||: | :||||||:|| |: ||||
Db          1 MILVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              :|||||||:||:||:|| |||||: | |:|||:|||||||||||||||||||||||||| |
 b         61 VISSVVPPMMFSLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRIVNAVAAIEL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || | |||||||||||| |:| ||| || ||||| ||||||| ||:|||||||  :|  ::
Db        121 YGYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQVV 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              | ||: :||||| :||| ||:|:||||| ||: : ||||||||| ||  ||: ||::| |
Db        181 GTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240

Qy        241 LTLKGLELIYERN 253
              ||||||:|||::|
Db        241 LTLKGLQLIYKKN 253


RESULT 3
Q8YAC5
ID   Q8YAC5      PRELIMINARY;      PRT;   259 AA.
AC   Q8YAC5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein lmo0221.
GN   LMO0221.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-E / SEROVAR 1/2A;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591974; CAD00748.1; -.
DR   MEROPS; M41.009; -.
DR   ListiList; LMO00221; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28187 MW;  F106B049D80025B7 CRC64;

  Query Match           63.8%;  Score 852;  DB 16;  Length 259;
  Best Local Similarity  63.4%;  Pred. No. 9.1e-68;
  Matches  161;  Conservative   37;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||||||||  :|||   ||  |||: | ||:| || || : : | :: |    | ||
Db          1 MILVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||||| |:| || :||:| | |||||:|||||:| |||:|:|:|||||||||
Db         61 IISSVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || |:||||||||||:|||||:  | |||||||| |||||||:|||||||::|    ||
Db        121 YGTPVIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||:|||:||:|| :|:||| |||:  || |:     |:|||||| :|  :|  :||:|||
Db        181 GKSTVSSMQAGIFYGFVGQCEGIIAEMKKQSNASPVVVATGGLARMITEKSSAVDILDPF 240

Qy        241 LTLKGLELIYERNR 254
              ||||||||:| ||:
Db        241 LTLKGLELLYRRNK 254


RESULT 4
Q92F54
ID   Q92F54      PRELIMINARY;      PRT;   259 AA.
AC   Q92F54;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein lin0253.
GN   LIN0253.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
```

```
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596164; CAC95486.1; -.
DR   ListiList; LIN00253; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28227 MW;  554B03A0C0EFA64F CRC64;

  Query Match             63.4%;  Score 847;  DB 16;  Length 259;
  Best Local Similarity   62.6%;  Pred. No. 2.5e-67;
  Matches  159;  Conservative   39;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||||||||  :||| : ||  |||: | ||:| || || : : | :: |    | ||
Db          1 MILVIDVGNTNCTVGVYKEQKLLRHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||||| |:| || :||:| | |||||:|||||:| |||:|:|:|||||||||
Db         61 IISSVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || |:||||||||||:|||||   | ||||||| |||||||:|||||||::|     ||
Db        121 YGTPVIVVDFGTATTFCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||:||::||:|| :|::|| |||:  || |:     |:|||||| :|  :|  :||:|||
Db        181 GKSTVASMQAGIFYGFIGQCEGIIAEMKKQSNTSPVVVATGGLARMITEKSSAVDILDPF 240

Qy        241 LTLKGLELIYERNR 254
              ||||||||:| ||:
Db        241 LTLKGLELLYRRNK 254


RESULT 5
Q97EB4
ID   Q97EB4      PRELIMINARY;      PRT;   273 AA.
AC   Q97EB4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Predicted transcriptional regulator, homolog of Bvg accessory
DE   factor.
GN   CAC3200.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007815; AAK81136.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30331 MW;  DE7B4D5923E72842 CRC64;

  Query Match             58.1%;  Score 775;  DB 16;  Length 273;
  Best Local Similarity   56.4%;  Pred. No. 6.9e-61;
  Matches  146;  Conservative   55;  Mismatches   54;  Indels    4;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||:|||||| |||:|:|: ||   ||: |   :: ||:|: : :||    |    ::|:
Db         12 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGV 71

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| ||::|| |  ||| | | :||||:|||:|||||||||||||||||||||  :
Db         72 IISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVAAHEI 131

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              |   ||::|||||||:| : ||  |:|||| ||| :|:|||: :||||||:|: :|   |
Db        132 YKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKPAYAI 191

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLK----VIATGGLAPLIANESDCIDI 236
               |||:|::||||::||:|||  ||:||| : ::: :    |:|||||| ||: |:  :|:
Db        192 CKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKLISEEAKNVDV 251

Qy        237 VDPFLTLKGLELIYERNRV 255
              ::||||||:|| :|||:|||
Db        252 INPFLTLEGLRIIYEKNRV 270


RESULT 6
Q8R7M2
ID   Q8R7M2      PRELIMINARY;      PRT;   255 AA.
AC   Q8R7M2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcriptional regulator, homologs of Bvg accessory
DE   factor.
GN   TTE2381.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013180; AAM25520.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   255 AA;  27816 MW;  C3C620ECBC8CA6ED CRC64;

  Query Match             57.9%;  Score 773;  DB 16;  Length 255;
  Best Local Similarity   57.1%;  Pred. No. 9.5e-61;
  Matches  145;  Conservative   48;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :||  |||||| |:||:   || : :|| | ::|| ||:||::  |  ::|:   :|| :
Db          1 MLLAFDVGNTNIVMGVFKGKKLLHSFRISTDKNKTYDEYGMLVNQLIGYNGISLTEIDDV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||:|  |: |  |||  :| :||||:|||:|||||||||||||||||||  |
Db         61 IISSVVPPLMNTLQVMSLKYFRTKPIVVGPGIKTGINIKYDNPKEVGADRIVNAVAAYEL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || |:||:|||||||:| | |  :|:|| |||| || :||: | ||||:|::|:|  :|
Db        121 YGGPVIVIDFGTATTFCAISEKGEYLGGIIAPGLMISADALFQRTAKLPKIDLTKPPTVI 180
```

```
Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              :|||::|||||::|:|| |: || ||| :    |:|||| | :|| ||  || |:
Db        181 NRNTVASMQSGIIYGHVGMVDYIVTRMKGEFAPSAYVVATGGFANMIAEESKTIDTVNEM 240

Qy        241 LTLKGLELIYERNR 254
              |||:|| :|||||:
Db        241 LTLEGLRIIYERNK 254


RESULT 7
Q8XHL5
ID   Q8XHL5      PRELIMINARY;      PRT;   259 AA.
AC   Q8XHL5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CPE2468.
GN   CPE2468.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003194; BAB82174.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28819 MW;  0D5FEA3B7A145E10 CRC64;

  Query Match             55.2%;  Score 737;  DB 16;  Length 259;
  Best Local Similarity   56.2%;  Pred. No. 1.5e-57;
  Matches  145;  Conservative   50;  Mismatches   59;  Indels    4;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|:|||||| |||:: : |    ||| |   || ||: : :  ||: : |  | ::||
Db          1 MILLIDVGNTNIVLGIHDNEKYIASWRISTDSKKTSDEYSIQVMQLFNQAKLNPEDVEGI 60

 y         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| || :|| |  | |  || :||||:|||:||||||||||||||||||||
Db         61 IISSVVPNIMHSLENMVRKCFCKEPIVVGPGIKTGINIKYDNPKEVGADRIVNAVAAFEK 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              :  |:|::|||||||:| | |   |:|| | ||| || :||: |||||||||: :| ::|
Db        121 HKKPMIIIDFGTATTFCAITEKGDYLGGNICPGIQISADALFERAAKLPRIELEKPKSVI 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQ----AKQDLKVIATGGLAPLIANESDCIDI 236
              ||||::||:||::||:|:|| |||||| :     :::  |:|||||| |: :|:| ||
Db        181 CKNTVTSMQAGIIYGYIGKVEYIVKRMKKEMMDLGEKEPFVLATGGLAKLVYSETDVIDE 240

Qy        237 VDPFLTLKGLELIYERNR 254
              ||  |||:||:::||:|:
Db        241 VDRKLTLEGLKILYEKNK 258


RESULT 8
Q9X8N6
ID   Q9X8N6      PRELIMINARY;      PRT;   265 AA.
AC   Q9X8N6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO3380.
GN   SCO3380 OR SCE94.31C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL049628; CAB40880.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   265 AA;  28203 MW;  E34FA2ABBA946969 CRC64;

  Query Match             49.1%;  Score 655;  DB 16;  Length 265;
  Best Local Similarity   51.0%;  Pred. No. 3.1e-50;
  Matches  133;  Conservative   46;  Mismatches   74;  Indels    8;  Gaps    3;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ---- 56
              :|| ||||||:||||::     :  |||| |   :| ||  ::|: |     |: ::
Db          1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDG 60

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              |||| | : || ::  |  :  :|:   | : | ||:|||: |  |:||||||||:|||
Db         61 IDGIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPILTDHPKEVGADRIINAV 120

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||: ||| | |||||||||||:  :    :|:|| ||||| || |||  : |:| :||: |
Db        121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVAR 180

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQA---KQDLKVIATGGLAPLIANESD 232
              | ::|||||| ||||||::|: |||:|:| ||  :      |: |||||||||::  ||
Db        181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGGLAPMVLGESS 240

Qy        233 CIDIVDPFLTLKGLELIYERN 253
              ||  :|:||| || |:||||
Db        241 VIDEHEPWLTLMGLRLVYERN 261


RESULT 9
```

```
Q9A6Z1
ID   Q9A6Z1      PRELIMINARY;      PRT;   261 AA.
AC   Q9A6Z1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcriptional activator, putative, Baf family.
GN   CC1935.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005867; AAK23910.1; -.
DR   TIGR; CC1935; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   261 AA;  27965 MW;  C19E60D7B0714EF5 CRC64;

  Query Match             43.2%;  Score 577;  DB 16;  Length 261;
  Best Local Similarity   44.4%;  Pred. No. 2.7e-43;
  Matches  114;  Conservative   50;  Mismatches   91;  Indels    2;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|| |: ||||||:  ::         ||  |   :| ||: : |  |    || |  || :
Db          2 MLLAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAV 61

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||  :| |  :  :||::|| ::|   | |:::: : | | ||||:|||: |  :
Db         62 IISSVVPQSIPNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 121

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRP--DN 178
              |  ||:|:| |||||:  :  :  : || ||||| :| :||:  |||||||  | ||  :
Db        122 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 181

Qy        179 IIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVD 238
              |:| :||||||||: :||:  :||:| |:|  :  : : |||||:| |    :| ||  |
Db        182 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFD 241

Qy        239 PFLTLKGLELIYERNRV 255
                ||::||  || || :
Db        242 SDLTIRGLLEIYRRNTI 258


RESULT 10
Q9CD56
ID   Q9CD56      PRELIMINARY;      PRT;   274 AA.
AC   Q9CD56;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein ML0232.
GN   ML0232.
OS   Mycobacterium leprae.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
DR   EMBL; AL583917; CAC29740.1; -.
DR   Leproma; ML0232; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   274 AA;  29421 MW;  1C2E735BDEC78765 CRC64;

  Query Match             38.9%;  Score 519.5;  DB 16;  Length 274;
  Best Local Similarity   41.0%;  Pred. No. 3.7e-38;
  Matches  109;  Conservative   53;  Mismatches   89;  Indels   15;  Gaps    5;

Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  :|:  |        |:
Db          1 MLLAIDVRNTHTVVGLLSGSKEHAKVVQQWRIRTESEVTADELALIIDGLIGDDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : : ||::||: :  ||||||||||||| :
Db         58 LAGAAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              || | :|   ||||||::      :    :::|||||||: :|::|  :|:| | |:|: |
Db        118 AAFHKFGQAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQ---DL----KVIATGGLAPLIA 228
              | :::|||||  ||:|::||: | |:|:| ||:   ::   ||     |:||| |||:
Db        178 PRSVVGKNTVECMQAGVVFGFAGLVDGLVGRMRQDVEEFSGDLGNRVAVVATGHTAPLLL 237

Qy        229 NESDCIDIVDPFLTLKGLELIYERNR 254
                |  :|  |  ||| || |::|||
Db        238 PELHTVDHYDRHLTLHGLRLVFERNR 263


RESULT 11
O06282
ID   O06282      PRELIMINARY;      PRT;   272 AA.
AC   O06282;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 29.3 kDa protein (Transcriptional activator, putative,
DE   Baf family).
GN   RV3600C OR MTCY07H7B.22 OR MT3706.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
```

```
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z95557; CAB08944.1; -.
DR   EMBL; AE007170; AAK48063.1; -.
DR   TIGR; MT3706; -.
DR   TubercuList; Rv3600c; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   272 AA;  29304 MW;  5D70E6EF0F09AC8B CRC64;

  Query Match             37.0%;  Score 494.5;  DB 16;  Length 272;
  Best Local Similarity   38.8%;  Pred. No. 6.2e-36;
  Matches  102;  Conservative   55;  Mismatches   93;  Indels   13;  Gaps    4;

Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  : :  |         |:
Db          1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : : ||::||: :  ||||||||||||| :
Db         58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||   :    ||||||::     :    :::|||||||: :|::|  :|:| | |:|: |
Db        118 AAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
              | :::|||||  ||:| :||: | |:|:| |::     :    |: ::|||  |||:  |
Db        178 PRSVVGKNTVECMQAGAVFGFAGLVDGLVGRIREDVSGFSVDHDVAIVATGHTAPLLLPE 237

Qy        231 SDCIDIVDPFLTLKGLELIYERN 253
                :|  |  |||:|| |::|||
Db        238 LHTVDHYDQHLTLQGLRLVFERN 260


RESULT 12
Q8RFE4
ID   Q8RFE4      PRELIMINARY;      PRT;   256 AA.
AC   Q8RFE4;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bvg accessory factor.
GN   FN0761.
OS   Fusobacterium nucleatum (subsp. nucleatum).
OC   Bacteria; Fusobacteria; Fusobacterium.
OX   NCBI_TaxID=76856;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 25586;
RX   MEDLINE=21886394; PubMed=11889109;
RA   Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA   Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA   Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA   Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA   Fonstein M., Kyrpides N., Overbeek R.;
RT   "Genome sequence and analysis of the oral bacterium Fusobacterium
RT   nucleatum strain ATCC 25586.";
RL   J. Bacteriol. 184:2005-2018(2002).
DR   EMBL; AE010586; AAL94957.1; -.
KW   Complete proteome.
SQ   SEQUENCE   256 AA;  28601 MW;  CDF681127F08914B CRC64;

  Query Match             34.3%;  Score 458;  DB 16;  Length 256;
  Best Local Similarity   38.1%;  Pred. No. 1e-32;
  Matches   98;  Conservative   60;  Mismatches   95;  Indels    4;  Gaps    4;

Qy          1 LLLVIDVGNTNTVLGVY-HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDG 59
              ::: ||:|||: | |:| ::|:|   :|| |:   ||||:      ::  :: :  :::|
Db          1 MIIGIDIGNTHIVTGIYDNNGELISTFRIATNDKMTEDEYFSYFNNITKYNEISIKKVDA 60

Qy         60 IIISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNI-KYDNPKEVGADRIVNAVAAI 118
              |:|||||| |:   :    ||| :|  ||    |      |  |    |||||::   |:
Db         61 ILISSVVPNIIITFQFFARKYFKVEATIVDLEKKLPFTFAKGINYTGFGADRIIDITEAM 120

Qy        119 HLYGNP-LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPD 177
                | :  |:: |||||||  : :   |:|| | ||| :|  |||   |||||:: | |
Db        121 QKYPDKNLVIFDFGTATTYDVLKKG-VYIGGGILPGIDMSINALYGNTAKLPRVKFTTPS 179

Qy        178 NIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIV 237
              :::| :|:  :|: | ||| ||:: |:|::  :  ::: |:|||||  ::: | | ||
Db        180 SVLGTDTMKQIQAAIFFGYAGQIKHIIKKINEELNEEIFVLATGGLGKILSAEIDEIDEY 239

Qy        238 DPFLTLKGLELIYERNR 254
              |  |:||||  :|: |:
Db        240 DANLSLKGLYTLYKLNK 256


RESULT 13
Q9RX54
ID   Q9RX54      PRELIMINARY;      PRT;   262 AA.
AC   Q9RX54;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein DR0461.
GN   DR0461.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE001905; AAF10040.1; -.
DR   TIGR; DR0461; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001220; Lectin_legB.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00307; LECTIN_LEGUME_BETA; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   262 AA;  27839 MW;  965EAAD2F78785A0 CRC64;

  Query Match             32.1%;  Score 428;  DB 16;  Length 262;
  Best Local Similarity   37.4%;  Pred. No. 4.9e-30;
  Matches   98;  Conservative   51;  Mismatches   99;  Indels   14;  Gaps    6;

Qy          2 LLVIDVGNTNTVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              || :|:||| ||||:    | | : ||| |:|    |:  : |  ||  :|    :
Db          6 LLAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAGAPIPR--AA 63
```

```
Qy          61 IISSVVPPI----MFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVA 116
               ::||| ||:      ||:|    :| |:   |       : :: | |  |||||: |
Db          64 VLSSVAPPVGENYALALKR----HFMIDAFAVSAENLPDVTVELDTPGSVGADRLCNLFG 119

Qy         117 AIHLYG--NPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEIT 174
               |    |  :  :|||||:| :  :   ::::|| :| |  :| :||::|||||||| :
Db         120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy         175 RPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCI 234
                |:  |||||| |:|||::|||   |:|:::|:: :   :   :|||| :  :      |
Db         180 APETAIGKNTVHALQSGLVFGYAEMVDGLLRRIRAELPGEAVAVATGGFSRTVQGICQEI 239

Qy         235 DIVDPFLTLKGL-ELIYERNRV 255
               |  |  |||:|| ||   |: |
Db         240 DYYDETLTLRGLVELWASRSEV 261


RESULT 14
Q9WZY5
ID   Q9WZY5      PRELIMINARY;      PRT;   246 AA.
AC   Q9WZY5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TM0883.
GN   TM0883.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001754; AAD35964.1; -.
DR   TIGR; TM0883; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   246 AA;  27154 MW;  9E0309AD462CF266 CRC64;

  Query Match             30.4%;  Score 405.5;  DB 16;  Length 246;
  Best Local Similarity   39.2%;  Pred. No. 4.5e-28;
  Matches  100;  Conservative   42;  Mismatches   88;  Indels   25;  Gaps    9;

Qy           1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDH----SGLMFEQ 56
               : |::|||||::|  :  |||    ||: |   :||||       || |     |    :
Db           1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-------LFSHLHPLLGDAMRE 53

Qy          57 IDGIIISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGL---NIKYDNPKEVGADRIVN 113
               | || ::|||      :||   ||||| |  :    | |    |:|  || ||||||: |
Db          54 IKGIGVASVVPTQNTVIERFSQKYFHISP--IWVKAKNGCVKWNVK--NPSEVGADRVAN 109

Qy         114 AVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEI 173
                || :  ||   |::| |||||   :  |  | |||| ||  :   :|:   |||| :|:
Db         110 VVAFVKEYGKNGIIIDMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168

Qy         174 TRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233
                  | ::||:|   :: |:: | |  :|||: |:| :   || |: ||| : ::    |
Db         169 KPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIK-EVYGDLPVVLTGGQSKIV---KDM 224

Qy         234 I--DIVDPFLTLKGL 246
               |  :| |  ||:||:
Db         225 IKHEIFDEDLTIKGV 239


RESULT 15
O32514
ID   O32514      PRELIMINARY;      PRT;   212 AA.
AC   O32514;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Simillar to Bacillus subtilis.
OS   Desulfovibrio vulgaris (strain Miyazaki).
OC   Bacteria; Proteobacteria; delta subdivision; Desulfovibrionaceae;
OC   Desulfovibrio.
OX   NCBI_TaxID=883;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MIYAZAKI;
RA   Kitamura M., Konishi T., Kawanishi K., Ohashi K., Kishida Y.,
RA   Kohno K., Akutsu H., Kumagai I., Nakaya T.;
RT   "Sequence analyses of two ferredoxin genes and their flanking regions
RT   from Desulfovibrio vulgaris (Miyazaki F).";
RL   J. Biochem. Mol. Biol. Biophys. 2:147-154(1998).
DR   EMBL; AB005550; BAA21476.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
SQ   SEQUENCE   212 AA;  22336 MW;  23E1789973A344D7 CRC64;

  Query Match             25.4%;  Score 338.5;  DB 2;  Length 212;
  Best Local Similarity   37.2%;  Pred. No. 3.4e-22;
  Matches   77;  Conservative   40;  Mismatches   83;  Indels    7;  Gaps    5;

Qy           3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62
               |: |:|||| :|:  :  :   : : |   :| |  |: |  :  |:||    :   :
Db           6 LLFDIGNTNVKIGIAVETAVLTSYVLPTDPGQTTDSIGLRLLEVLRHAGLGPADVGACVA 65

Qy          63 SSVVPPIMFALERMCTKYFHIEPQIVGPG-MKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121
               ||||| :   : | | :| :    :  || :   |: :|: | ||||||:| | ||  ||
Db          66 SSVVPGVNPLIRRACERYLY-RKLLFAPGDIAIPLDNRYERPAEVGADRLVAAYAARRLY 124

Qy         122 GNP--LIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRI--EITRPD 177
                 |  |: ||||||||:  : |   |:|| | ||:  |  || || ||||||  |:
Db         125 PGPRSLVSVDFGTATTFDCV-EGGAYLGGLICPGVLSSAGALSSRTAKLPRISLEVEEDS 183

Qy         178 NIIGKNTVSAMQSGILFGYVGQVEGIV 204
               :|||:| :::  | :||:    ||::
Db         184 PVIGRSTTTSLNHGFIFGFAAMTEGVL 210


Search completed: June 24, 2003, 21:59:13
Job time : 24.0426 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.48734 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1114	83.4	233	1	YACB_BACSU	P37564 bacillus su
2	129.5	9.7	267	1	BAF_BORPE	Q45338 bordetella
3	90.5	6.8	4451	1	GRSB_BACBR	P14688 b gramicidi
4	89.5	6.7	357	1	BUK_THETN	Q8r832 thermoanaer
5	88	6.6	256	1	HI61_PSEAE	Q9hu44 pseudomonas
6	87.5	6.6	424	1	SAHH_METKA	P58855 methanopyru
7	87	6.5	336	1	PYRD_SALTY	P25468 salmonella
8	86.5	6.5	243	1	Y004_METJA	Q60315 methanococc
9	85	6.4	336	1	PYRD_ECOLI	P05021 escherichia
10	83.5	6.3	204	1	DMSB_ECOLI	P18776 escherichia
11	83	6.2	300	1	YHCI_HAEIN	P44541 haemophilus
12	83	6.2	425	1	GSA_CLOPE	Q9znc8 clostridium
13	83	6.2	470	1	CH60_TRIVA	Q95058 trichomonas
14	82	6.1	305	1	GALE_METJA	Q57664 methanococc
15	82	6.1	3898	1	POLG_HCVB	P21530 hog cholera
16	81	6.1	359	1	OP21_HAEIN	P43839 haemophilus
17	81	6.1	1224	1	RPB2_YEAST	P08518 saccharomyc
18	80.5	6.0	208	1	PIMT_YERPE	Q8zbq0 yersinia pe
19	80.5	6.0	327	1	EBGR_ECOLI	P06846 escherichia
20	80.5	6.0	354	1	LUXA_VIBFI	P19907 vibrio fisc
21	80.5	6.0	410	1	RAA2_CHLRE	Q9smh4 chlamydomon
22	80	6.0	449	1	MURD_STAAM	O33595 staphylococ
23	79.5	6.0	346	1	GCP_BACSU	O05518 bacillus su
24	79	5.9	304	1	HEM3_AQUAE	O66621 aquifex aeo
25	79	5.9	314	1	Y198_METJA	Q57651 methanococc
26	79	5.9	361	1	OP22_HAEIN	P20149 haemophilus
27	79	5.9	923	1	NRP1_MOUSE	P97333 mus musculu
28	79	5.9	991	1	SCA4_RICAU	Q9aj64 rickettsia
29	78.5	5.9	337	1	GRA4_MOUSE	Q61603 mus musculu
30	78.5	5.9	453	1	TBB2_GEOCN	P32925 geotrichum
31	78.5	5.9	505	1	ALAT_SCHPO	Q10334 schizosacch
32	78	5.8	208	1	PIMT_SALTY	Q8zmf9 salmonella
33	78	5.8	469	1	LEU2_AGRT5	Q8uby9 agrobacteri
34	78	5.8	1032	1	A1A4_MOUSE	Q9wv27 mus musculu
35	78	5.8	1067	1	CARB_CLOPE	Q8xhb3 clostridium
36	78	5.8	1264	1	MOLR_ECOLI	P33345 escherichia
37	78	5.8	1342	1	RPOB_BUCAI	P57146 buchnera ap
38	78	5.8	3898	1	POLG_HCVA	P19712 hog cholera
39	77.5	5.8	204	1	YNFG_ECOLI	P77313 escherichia
40	77.5	5.8	291	1	PYRD_SULAC	O08358 sulfolobus
41	77.5	5.8	352	1	EGSA_HALN1	Q9hs49 halobacteri
42	77.5	5.8	355	1	BUK_LISIN	Q92by5 listeria in
43	77.5	5.8	393	1	AAT1_BACSU	P53001 bacillus su
44	77.5	5.8	428	1	PURA_CLOAB	Q97d87 clostridium
45	77.5	5.8	484	1	HXK1_SCHPO	Q09756 schizosacch

ALIGNMENTS

```
RESULT 1
YACB_BACSU
ID   YACB_BACSU     STANDARD;      PRT;   233 AA.
AC   P37564;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yacB.
GN   YACB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match           83.4%;  Score 1114;  DB 1;  Length 233;
  Best Local Similarity   99.5%;  Pred. No. 8.9e-86;
  Matches  213;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214
              |||||||||||||||||||||||||||||||||
Db        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214


RESULT 2
BAF_BORPE
ID   BAF_BORPE      STANDARD;      PRT;   267 AA.
AC   Q45338; Q45373;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Bvg accessory factor.
GN   BAF.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RX   MEDLINE=95325323; PubMed=7601846;
RA   Deshazer D., Wood G.E., Friedman R.L.;
RT   "Identification of a Bordetella pertussis regulatory factor required
RT   for transcription of the pertussis toxin operon in Escherichia
RT   coli.";
RL   J. Bacteriol. 177:3801-3807(1995).
RN   [2]
RP   SEQUENCE OF 1-38 FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 239-267 FROM N.A.
RC   STRAIN=BP536;
RX   MEDLINE=96419162; PubMed=8821935;
RA   Allen A.G., Maskell D.J.;
RT   "The identification, cloning and mutagenesis of a genetic locus
RT   required for lipopolysaccharide biosynthesis in Bordetella
RT   pertussis.";
RL   Mol. Microbiol. 19:37-52(1996).
CC   -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC       A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC       RNA POLYMERASE.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U12020; AAA75361.1; -.
DR   EMBL; AF016461; AAC68834.1; -.
DR   EMBL; X90711; CAA62242.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Transcription regulation; Activator.
SQ   SEQUENCE   267 AA;  27845 MW;  336A615F67B57901 CRC64;

  Query Match           9.7%;  Score 129.5;  DB 1;  Length 267;
  Best Local Similarity   31.0%;  Pred. No. 0.00093;
  Matches   49;  Conservative   22;  Mismatches   68;  Indels   19;  Gaps    6;

Qy         83 IEPQIVGPGMKTGLNIKYDNPKEVGADR---IVNAVAAIHLYGNPLIVVDFGTATTYCYI 139
              :  | :  |:: |    | || ::|||   :|  :|       ||:|  ||||||  |
Db         83 LRAQPLAMGLRNG----YRNPDQLGADRWACMVGVLARQPSVHPPLLVASFGTATTLDTI 138

Qy        140 DENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNTVSAMQSGILFGYVGQ 199
              :  : || | ||  :   ||    | ||  :    |  |  :|  |: |||     |
Db        139 GPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPI--DTHQAIASGIA---AAQ 193

Qy        200 VEGIVKRMKWQA-----KQDLKVIATGGLAPLIANESD 232
                ||:  :| |      |  ::   ||  | :  |::
Db        194 AGAIVR--QWLAGRQRYGQAPEIYVAGGGWPEVRQEAE 229


RESULT 3
GRSB_BACBR
ID   GRSB_BACBR     STANDARD;      PRT;  4451 AA.
AC   P14688;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Gramicidin S synthetase II [Includes: ATP-dependent proline adenylase
DE   (ProA) (Proline activase); ATP-dependent valine adenylase (ValA)
DE   (Valine activase); ATP-dependent ornithine adenylase (OrnA) (Ornithine
DE   activase); ATP-dependent leucine adenylase (LeuA) (Leucine activase)].
GN   GRSB OR GRS2.
OS   Bacillus brevis.
OC   Bacteria; Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
OX   NCBI_TaxID=1393;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 9999;
RX   MEDLINE=92219998; PubMed=1560782;
RA   Turgay K., Krause M., Marahiel M.A.;
RT   "Four homologous domains in the primary structure of GrsB are related
RT   to domains in a superfamily of adenylate-forming enzymes.";
RL   Mol. Microbiol. 6:529-546(1992).
RN   [2]
RP   SEQUENCE OF 1-948 FROM N.A.
RC   STRAIN=Nagano;
RX   MEDLINE=92041751; PubMed=1939016;
RA   Hori K., Yamamoto Y., Tokita K., Saito F., Kurotsu T., Kanda M.,
RA   Okamura K., Furuyama J., Saito Y.;
RT   "The nucleotide sequence for a proline-activating domain of
RT   gramicidin S synthetase 2 gene from Bacillus brevis.";
RL   J. Biochem. 110:111-119(1991).
RN   [3]
```

```
RP   SEQUENCE OF 1-143 FROM N.A.
RC   STRAIN=ATCC 9999;
RX   MEDLINE=90008776; PubMed=2477357;
RA   Kraetzschmar J., Krause M., Marahiel M.A.;
RT   "Gramicidin S biosynthesis operon containing the structural genes
RT   grsA and grsB has an open reading frame encoding a protein homologous
RT   to fatty acid thioesterases.";
RL   J. Bacteriol. 171:5422-5429(1989).
RN   [4]
RP   SEQUENCE OF 1-15, AND CHARACTERIZATION.
RC   STRAIN=Nagano;
RX   MEDLINE=92011463; PubMed=1917901;
RA   Kurotsu T., Hori K., Kanda M., Saito Y.;
RT   "Characterization and location of the L-proline activating fragment
RT   from the multifunctional gramicidin S synthetase 2.";
RL   J. Biochem. 109:763-769(1991).
CC   -!- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME, ABLE TO
CC       ACTIVATE AND POLYMERIZE THE AMINO ACIDS PRO, VAL, ORN AND LEU.
CC       ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
CC   -!- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF THE CYCLIC PEPTIDE
CC       ANTIBIOTIC GRAMICIDIN S (D-PHE-PRO-VAL-ORN-LEU)2.
CC   -!- SUBUNIT: LARGE MULTIENZYME COMPLEX OF GRSA AND GRSB.
CC   -!- DOMAIN: CONSISTS OF FOUR MODULES, AND HARBORS A PUTATIVE
CC       THIOESTERASE DOMAIN AT ITS C-TERMINAL END. EACH MODULE
CC       INCORPORATES ONE AMINO ACID INTO THE PEPTIDE PRODUCT AND CAN BE
CC       FURTHER SUBDIVIDED INTO DOMAINS RESPONSIBLE FOR SUBSTRATE
CC       ADENYLATION, THIOLATION, CONDENSATION (NOT FOR THE INITIATION
CC       MODULE), AND EPIMERIZATION (OPTIONAL), AND N METHYLATION
CC       (OPTIONAL).
CC   -!- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC       FAMILY.
CC   -!- SIMILARITY: CONTAINS 4 ACYL CARRIER DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X61658; CAA43838.1; -.
DR   EMBL; M29703; AAA58719.1; -.
DR   EMBL; X15577; CAA33604.1; -.
DR   EMBL; D00938; BAA00778.1; -.
DR   PIR; S20542; YGBSG2.
DR   HSSP; P14687; 1AMU.
DR   InterPro; IPR000873; AMP-bind.
DR   InterPro; IPR001242; Condensatn.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR000379; Ser_estrs_site.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00501; AMP-binding; 4.
DR   Pfam; PF00550; pp-binding; 4.
DR   Pfam; PF00668; Condensation; 4.
DR   Pfam; PF00975; Thioesterase; 1.
DR   PRINTS; PR00154; AMPBINDING.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 4.
DR   PROSITE; PS00455; AMP_BINDING; 4.
DR   PROSITE; PS50075; ACP_DOMAIN; 4.
KW   Ligase; Antibiotic biosynthesis; Multifunctional enzyme;
KW   Repeat; Phosphopantetheine; Hydrolase.
FT   INIT_MET      0      0
FT   REPEAT      466   1043       DOMAIN 1 (PROLINE-ACTIVATING).
FT   REPEAT     1521   2080       DOMAIN 2 (VALINE-ACTIVATING).
FT   REPEAT     2538   3135       DOMAIN 3 (ORNITHINE-ACTIVATING).
FT   REPEAT     3591   4173       DOMAIN 4 (LEUCINE-ACTIVATING).
FT   DOMAIN      975   1042       ACYL CARRIER (ACP) 1.
FT   DOMAIN     2011   2078       ACYL CARRIER (ACP) 2.
FT   DOMAIN     3057   3124       ACYL CARRIER (ACP) 3.
FT   DOMAIN     4095   4162       ACYL CARRIER (ACP) 4.
FT   BINDING    1005   1005       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    2041   2041       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    3087   3087       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   BINDING    4125   4125       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT    274    274       H -> D (IN REF. 2).
FT   CONFLICT    418    418       K -> R (IN REF. 2).
FT   CONFLICT    654    664       NAVLTCVTKKF -> TCSFDVCYQEI (IN REF. 2).
FT   CONFLICT    941    946       HVRLHL -> QLPLTP (IN REF. 2).
SQ   SEQUENCE   4451 AA;  510036 MW;  E0029C9B51F5A4B7 CRC64;

  Query Match             6.8%;  Score 90.5;  DB 1;  Length 4451;
  Best Local Similarity   22.7%;  Pred. No. 44;
  Matches   56;  Conservative   35;  Mismatches   77;  Indels   79;  Gaps   14;

Qy         70 MFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVG----------------ADRIV 112
              || |:    | | :|   : | :    : |:|   ||                  |: |
Db       1396 MFVLQNTDRKSFEVEQITITPYVPNSRHSKFDLTLEVSEEQNEILLCLEYCTKLFTDKTV 1455

Qy        113 NAVAA-----IH-LYGNPLIVV---------------DFG-TATTYCYIDENKQYM--- 146
                :|      :| : ||| |::               :|  | |||    : | |:
Db       1456 ERMAGHFLQILHAIVGNPTIIISEIEILSEEEKQHILFEFNDTKTTYPLCKQFKDYLRNR 1515

Qy        147 ----GGAIAPG---ITISTEALYSRAAKLPRI---EITRPDNIIGKNTVSAMQSGILFGY 196
                   :| |    |::   |  || :: |:   :  :||||:|  :     :| |
Db       1516 WRRRADHVAVGWKDQTLTYRELNERANQVARVLRQKGVQPDNIVG--LLVERSPEMLVG- 1572

Qy        197 VGQVEGIVK--------------RMKWQAKQDLKV---IATGGLAPLIANESDCIDIVD 238
                 : ||:|               |: :   ||  |   :    |  |: :| ||: |:|
Db       1573 ---IMGILKAGGAYLPLDPEYPADRISYMI-QDCGVRIMLTQQHLLSLVHDEFDCV-ILD 1627

Qy        239 PFLTLKG 245
                 ||
Db       1628 EDSLYKG 1634


RESULT 4
BUK_THETN
ID   BUK_THETN      STANDARD;      PRT;   357 AA.
AC   Q8R832;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable butyrate kinase (EC 2.7.2.7) (BK) (Branched-chain carboxylic
DE   acid kinase).
GN   BUK OR TTE2201.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC   Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4 / JCM 11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
CC   -!- CATALYTIC ACTIVITY: ATP + 2-butanoate = ADP + butanoyl phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ACETOKINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE013165; AAM25356.1; -.
DR   PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR   PROSITE; PS01076; ACETATE_KINASE_2; 1.
KW   Transferase; Kinase; Complete proteome.
```

```
SQ   SEQUENCE   357 AA;  39217 MW;  05CDDE09A2E7EF3F CRC64;

  Query Match             6.7%;  Score 89.5;  DB 1;  Length 357;
  Best Local Similarity   19.9%;  Pred. No. 2.8;
  Matches   66;  Conservative   43;  Mismatches   94;  Indels  129;  Gaps   16;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTE---------DEFGMILRSLFD--- 48
              |:|||: |:|:|: : |: | :  :    || || ||         |:|    ::: |
Db          5 LILVINPGSTSTKVAVFRDKEPVF---TETLRHSTEELSKYKRIIDQFEFRTQAILDMLK 61

Qy         49 HSGLMFEQIDGII--------ISS---------------------------------- 64
                |:   ||| |:          | |
Db         62 EKGISLSQIDAIVGRGGLLKPIESGTYIVNEKMLEDLKKAERGEHASNLGAIIAYTLAKE 121

Qy         65 -------VVPPIMFALERMC----------TKYFH-IEPQIVGPGMKTGLNIKYDNPKEV 106
                   | | ::  || :              || :  : :   : : |  :||
Db        122 HNIPAYIVDPVVVDELEDVARITGLPEIEKQSIFHALNQKAIARRLASDLGKRYDE---- 177

Qy        107 GADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMG-GAIAPGITISTEALYSRA 165
                  || : | || |      :  |       || |    | |  :|           ||
Db        178 -----VNLIIA-HLGGG----ISVGAHRKGRVIDVNDALNGEGPFSP----------ERA 217

Qy        166 AKLPRIEITRPDNIIGKNTVSAMQ-----SGILFGYVG----------------QVEGIV 204
                || ::: :     || |   |:       | :  ::|                  | |:
Db        218 GGLPVLDLVKL-CYSGKYTFEEMKKKLIGKGGIVAHLGTNDVREVYKMIENGDKNAELIL 276

Qy        205 KRMKWQAKQDLKVIATGGLAPLIANESDCIDI 236
                | :|  :::     | :| ::  : | | |
Db        277 DAMAYQTAKEI-----GSMAVVLKGKVDAIGI 303


RESULT 5
HI61_PSEAE
ID   HI61_PSEAE     STANDARD;      PRT;   256 AA.
AC   Q9HU44;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Imidazole glycerol phosphate synthase subunit hisF1 (EC 4.1.3.-) (IGP
DE   synthase cyclase subunit) (IGP synthase subunit hisF1) (ImGP synthase
DE   subunit hisF1) (IGPS subunit hisF1).
GN   HISF1 OR PA5140.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
CC   -!- FUNCTION: IGPS catalyzes the conversion of PRFAR and glutamine to
CC       IGP, AICAR and glutamate. The hisF subunit catalyzes the
CC       cyclization activity that produces IGP and AICAR from PRFAR using
CC       the ammonia provided by the hisH subunit (By similarity).
CC   -!- CATALYTIC ACTIVITY: 5-[(5-phospho-1-deoxyribulos-1-
CC       ylamino)methylideneamino]-1-(5-phosphoribosyl)imidazole-4-
CC       carboxamide + L-glutamine = imidazole-glycerol phosphate + 5-
CC       aminoimidazol-4-carboxamide ribonucleotide + L-glutamate + H(2)O.
CC   -!- PATHWAY: Histidine biosynthesis; fifth step.
CC   -!- SUBUNIT: Heterodimer of hisH and hisF (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE HISA / HISF FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004927; AAG08525.1; -.
DR   InterPro; IPR003009; FMN_enzyme.
DR   InterPro; IPR000570; His_biosynth.
DR   Pfam; PF00977; His_biosynth; 1.
DR   TIGRFAMs; TIGR00735; hisF; 1.
KW   Histidine biosynthesis; Lyase; Complete proteome.
FT   ACT_SITE     12     12       POTENTIAL.
FT   ACT_SITE    131    131       POTENTIAL.
SQ   SEQUENCE   256 AA;  27131 MW;  6486A76CD308FAA7 CRC64;

  Query Match             6.6%;  Score 88;  DB 1;  Length 256;
  Best Local Similarity   22.7%;  Pred. No. 2.5;
  Matches   60;  Conservative   45;  Mismatches   79;  Indels   80;  Gaps   15;

Qy          5 IDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIIISS 64
              :|| |   | ||  :   :    :| :|                :|  |   ::|  : |::
Db         11 LDVDNGRVVKGVKFENIRDAGDPVEIARR------------YDEQGA--DEITFLDITA 55

Qy         65 VV---PPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121
               |      :  :||| :: |   |  || |:::  :|:  |    |||::    ||:
Db         56 SVDGRDTTLHTVERMASQVF--IPLTVGGGVRSVQDIR--NLLNAGADKVSINTAAVF-- 109

Qy        122 GNPLIVVD----FGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEI---- 173
               ||  | :    ||:       ||  |       |||              : || ||
Db        110 -NPEFVGEAADRFGSQCIVVAIDAKK-----VSAPG-------------EAPRWEIFTHG 150

Qy        174 ----TRPDNII----------GKNTVSAM-QSGILFGY-VGQVEGIVKRMKWQAKQDLKV 217
                   |  | ::          |:  :::| | |:  || :|    | : :      :: |
Db        151 GRKPTGLDAVLWAKKMEDLGAGEILLTSMDQDGVKSGYDLGVTRAISEAV------NVPV 204

Qy        218 IATGGL-------APLIANESDCI 234
              ||:||:     ·  | ::  ::| :
Db        205 IASGGVGNLEHLAAGILEGKADAV 228


RESULT 6
SAHH_METKA
ID   SAHH_METKA     STANDARD;      PRT;   424 AA.
AC   P58855;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE   hydrolase) (AdoHcyase).
GN   AHCY OR MK0368.
OS   Methanopyrus kandleri.
OC   Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC   Methanopyrus.
OX   NCBI_TaxID=2320;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AV19 / DSM 6324 / JCM 9639;
RX   MEDLINE=21927647; PubMed=11930014;
RA   Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,
RA   Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA   Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA   Malykh A.G., Koonin E.V., Kozyavkin S.A.;
RT   "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT   and monophyly of archaeal methanogens.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC       + L-homocysteine.
CC   -!- COFACTOR: NAD (By similarity).
CC   -!- PATHWAY: Activated methyl cycle.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
```

```
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE010334; AAM01583.1; -.
DR   PROSITE; PS00738; ADOHCYASE_1; FALSE_NEG.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism; Complete proteome.
FT   NP_BIND     213    244       NAD (POTENTIAL).
SQ   SEQUENCE   424 AA;  47015 MW;  3DE7B10A3B44E1D5 CRC64;

  Query Match             6.6%;  Score 87.5;  DB 1;  Length 424;
  Best Local Similarity   20.8%;  Pred. No. 5;
  Matches   62;  Conservative   47;  Mismatches   96;  Indels   93;  Gaps   14;

Qy         19 DGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIIISSVVPPIMF-----AL 73
              :|   | || |     ||:|:              :: || ::  |  | |:       :
Db         97 EGVHVYAWRGE-----TEEEY-------------YQNIDRVL--SHEPDIIVDDGADCI 135

Qy         74 ERMCTKYFHIEPQIVGPGMK--TGLN----------------------IKY--DNPKEVG 107
               |: |::  :  :::|   :  ||:|                     ||  ||    |
Db        136 ARVHTEFPDLAERVIGATEETTTGVNRLHAMHREGVLKFPVIAVNDAKTKYLMDNRYGTG 195

Qy        108 ADRIVNAVAA--IHLYGNPLIVVDFGTATTYCYIDENKQYMG-GAIAPGITI----STEA 160
              :   : |  | | |  ::|| :|    :|     ::  | ||    : :    : ||
Db        196 QSALDGLMRATNILLAGKTVVVVGYG----WCGRGIARRARGLGANVIVVEVDPIKAMEA 251

Qy        161 LY--------SRAAKLPRIEITRPDN--IIGKNTVSAMQSGILFGYVGQVE--------- 201
              ::         |||:   | ||   |  :|    :  |: |::    |  :
Db        252 IFDGFRVMPMDRAAEEGDIFITATGNRDVIRGEHIEKMKDGVILANAGHFDVEIDKEYLE 311

Qy        202 ----------GIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPFLTLKGLEL 248
                       |:|   :    : : :|| | |  | | |   |:|:|    |: | :
Db        312 EHCEEKIDRRGGLVTEYRMPDGKRVYLIAEGRLVNLAAGEGHPIEIMDISFALQALSV 369


RESULT 7
PYRD_SALTY
ID   PYRD_SALTY     STANDARD;      PRT;   336 AA.
AC   P25468;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
DE   (DHOdehase) (DHODase) (DHOD).
GN   PYRD OR STM1058.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=91099333; PubMed=2269282;
RA   Frick M.M., Neuhard J., Kelln R.A.;
RT   "Cloning, nucleotide sequence and regulation of the Salmonella
RT   typhimurium pyrD gene encoding dihydroorotate dehydrogenase.";
RL   Eur. J. Biochem. 194:573-578(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC       H(2)O(2).
CC   -!- COFACTOR: FMN (BY SIMILARITY).
CC   -!- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.
CC       SUBFAMILY 2.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X55636; CAA39161.1; -.
DR   EMBL; AE008746; AAL19991.1; -.
DR   PIR; S13824; S13824.
DR   StyGene; SG10326; pyrD.
DR   InterPro; IPR001295; DHO_dh.
DR   InterPro; IPR003009; FMN_enzyme.
DR   Pfam; PF01180; DHOdehase; 1.
DR   TIGRFAMs; TIGR01036; pyrD_sub2; 1.
DR   PROSITE; PS00911; DHODEHASE_1; 1.
DR   PROSITE; PS00912; DHODEHASE_2; 1.
KW   Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KW   Complete proteome.
FT   NP_BIND     290    298       FMN (POTENTIAL).
SQ   SEQUENCE   336 AA;  36739 MW;  783F59D28DEE6398 CRC64;

  Query Match             6.5%;  Score 87;  DB 1;  Length 336;
  Best Local Similarity   21.6%;  Pred. No. 4.2;
  Matches   58;  Conservative   36;  Mismatches   79;  Indels   96;  Gaps   13;

Qy         37 DEFGMILRSLFDHSGL--MFEQI-----DGIIISSVVPPIMFALERMCTKYFHIEPQIVG 89
              |  |:| |  |:: |:  : | :     |||:
Db        105 DAEGLINRMGFNNLGVDNLVENVKKAHFDGIL---------------------------- 136

Qy         90 PGMKTGLNI--KYDNPKEVGADRIVNAVAAIHLYGNPLIV------------VDFGTATT 135
                   |:||    | | | | |  :  :  :: |   : :             : :| |
Db        137 -----GINIGKNKDTPVENGKDDYLICMEKVYAYAGYIAINISSPNTPGLRTLQYGDA-- 189

Qy        136 YCYIDE------NKQ---------YMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
                 :|:      |||          |:  |:     :  | |   |  | | |   | :|
Db        190 ---LDDLLTAIKNKQNDLQAIHHKYVPVAVKIAPDLCEEELIQVADSLLRHNI---DGVI 243

Qy        181 GKNTV----------SAMQSGILFGYVGQVEG--IVKRMKWQAKQDLKVIATGGLAPLIA 228
                ||           :  |:| | |   |::   |::|:  : |  | :|  ||:  .:||
Db        244 ATNTTLDRSLVQGMKNCQQTGGLSGRPLQLKSTEIIRRLSQELKGQLPIIGVGGIDSVIA 303

Qy        229 NE------SDCIDIVDPFLTLKGLELIYE 251
                    :  : |   |:  ||  || |
Db        304 AREKIAAGATLVQIYSGFI-FKGPPLIKE 331


RESULT 8
Y004_METJA
ID   Y004_METJA     STANDARD;      PRT;   243 AA.
AC   Q60315;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein MJ0004.
GN   MJ0004.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- SIMILARITY: A.FERMENTANS ACTIVATOR OF (R)-HYDROXYGLUTARYL-COA
CC       DEHYDRATASE (HGDC), TO E.COLI YJIL AND M.JANNASCHII MJ0800.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U67459; AAB97985.1; -.
DR   HSSP; P11568; 1HUX.
DR   TIGR; MJ0004; -.
DR   InterPro; IPR002731; ATPase_BadF.
DR   Pfam; PF01869; BcrAD_BadFG; 1.
DR   ProDom; PD006344; ATPase_BadF; 1.
DR   TIGRFAMs; TIGR00241; CoA_E_activ; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   243 AA;  26897 MW;  288C4F4190575E53 CRC64;

  Query Match             6.5%;  Score 86.5;  DB 1;  Length 243;
  Best Local Similarity   21.3%;  Pred. No. 3.2;
  Matches   54;  Conservative   44;  Mismatches   81;  Indels   75;  Gaps   12;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::| ||||:| | : :  | |: :       :| ||                    |:
Db          1 MILGIDVGSTTTKMVLMEDSKIIW--------YKIED-------------------IGV 32

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAI-- 118
              :|      |: : :   : : |: :||  |        |   |   ||:||  | |:
Db         33 VIEE---DILLKMVKEIEQKYPID-KIVATG--------YGRHKVSFADKIVPEVIALGK 80

Qy        119 ---HLYGNPLIVVDF-GTATTYCYIDENKQYMGGAIAPGITISTEALYSRA---AKLPRI 171
                 : :     |:|  |  |    ||:| : :   ::      |     :|    |: :
Db         81 GANYFFNEADGVIDIGGQDTKVLKIDKNGKVVDFILSDKCAAGTGKFLEKALDILKIDKN 140

Qy        172 EIT--RPDNIIGKNTVSAM--------------QSGILFG----YVGQVEGIVKRMKWQ 210
              ||   : |||   ::: |:               : ||| |     : :|  :  |:| |
Db        141 EINKYKSDNIAKISSMCAVFAESEIISLLSKKVPKEGILMGVYESIINRVIPMTNRLKIQ 200

Qy        211 AKQDLKVIATGGLA 224
                    :: :||:|
Db        201 -----NIVFSGGVA 209


RESULT 9
PYRD_ECOLI
ID   PYRD_ECOLI     STANDARD;      PRT;   336 AA.
AC   P05021;
DT   13-AUG-1987 (Rel. 05, Created)
DT   13-AUG-1987 (Rel. 05, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase)
DE   (DHOdehase) (DHODase) (DHOD).
GN   PYRD OR B0945 OR Z1294 OR ECS1029.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.
RX   MEDLINE=85285014; PubMed=2992959;
RA   Larsen N.J., Jensen K.F.;
RT   "Nucleotide sequence of the pyrD gene of Escherichia coli and
RT   characterization of the flavoprotein dihydroorotate dehydrogenase.";
RL   Eur. J. Biochem. 151:59-65(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC       H(2)O(2).
CC   -!- COFACTOR: FMN.
CC   -!- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: INNER SIDE OF THE MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE DIHYDROOROTATE DEHYDROGENASE FAMILY.
CC       SUBFAMILY 2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X02826; CAA26594.1; -.
```

```
DR   EMBL; AE000196; AAC74031.1; -.
DR   EMBL; D90732; BAA35700.1; -.
DR   EMBL; AE005284; AAG55431.1; -.
DR   EMBL; AP002553; BAB34452.1; -.
DR   PIR; A23109; DEECDO.
DR   SWISS-2DPAGE; P05021; COLI.
DR   EcoGene; EG10807; pyrD.
DR   InterPro; IPR001295; DHO_dh.
DR   InterPro; IPR003009; FMN_enzyme.
DR   Pfam; PF01180; DHOdehase; 1.
DR   TIGRFAMs; TIGR01036; pyrD_sub2; 1.
DR   PROSITE; PS00911; DHODEHASE_1; 1.
DR   PROSITE; PS00912; DHODEHASE_2; 1.
KW   Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FMN;
KW   Complete proteome.
FT   NP_BIND     290    298       FMN (POTENTIAL).
SQ   SEQUENCE   336 AA;  36774 MW;  973227EAE6B83622 CRC64;

  Query Match             6.4%;  Score 85;  DB 1;  Length 336;
  Best Local Similarity   25.0%;  Pred. No. 6.2;
  Matches   51;  Conservative   27;  Mismatches   70;  Indels   56;  Gaps   11;

Qy         95 GLNI--KYDNPKEVGADRIVNAVAAIHLYGNPLIV------------VDFGTATTYCYID 140
              |:||    | | | | |  :  :  |: |   : :            : :| |     :|
Db        137 GINIGKNKDTPVEQGKDDYLICMEKIYAYAGYIAINISSPNTPGLRTLQYGEA-----LD 191

Qy        141 E------NKQ---------YMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNT- 184
              :      |||         |:  |:     :| | |   |  | |  |   | :|  ||
Db        192 DLLTAIKNKQNDLQAMHHKYVPIAVKIAPDLSEEELIQVADSLVRHNI---DGVIATNTT 248

Qy        185 -----VSAM----QSGILFGYVGQVEG--IVKRMKWQAKQDLKVIATGGLAPLIANE--- 230
                  |  |    |:| | |   |::   |::|:  :    | :|  ||:  :||
Db        249 LDRSLVQGMKNCDQTGGLSGRPLQLKSTEIIRRLSLELNGRLPIIGVGGIDSVIAAREKI 308

Qy        231 ---SDCIDIVDPFLTLKGLELIYE 251
                :  : |   |:  ||  || | .
Db        309 AAGASLVQIYSGFI-FKGPPLIKE 331


RESULT 10
DMSB_ECOLI
ID   DMSB_ECOLI     STANDARD;      PRT;   204 AA.
AC   P18776; P77745;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Anaerobic dimethyl sulfoxide reductase chain B (DMSO reductase iron-
DE   sulfur subunit).
GN   DMSB OR B0895.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.
RC   STRAIN=K12 / C600;
RX   MEDLINE=89096500; PubMed=3062312;
RA   Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT   "Nucleotide sequence of the dmsABC operon encoding the anaerobic
RT   dimethylsulphoxide reductase of Escherichia coli.";
RL   Mol. Microbiol. 2:785-795(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97061202; PubMed=8905232;
RA   Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA   Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA   Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA   Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA   Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA   Yano M., Horiuchi T.;
RT   "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT   corresponding to the 12.7-28.0 min region on the linkage map.";
RL   DNA Res. 3:137-155(1996).
RN   [4]
RP   EPR SPECTROSCOPY OF IRON-SULFUR CLUSTERS.
RX   MEDLINE=91070067; PubMed=2174699;
RA   Cammack R., Weiner J.H.;
RT   "Electron paramagnetic resonance spectroscopic characterization of
RT   dimethyl sulfoxide reductase of Escherichia coli.";
RL   Biochemistry 29:8410-8416(1990).
RN   [5]
RP   MUTAGENESIS.
RX   MEDLINE=91355180; PubMed=1653010;
RA   Rothery R.A., Weiner J.H.;
RT   "Alteration of the iron-sulfur cluster composition of Escherichia
RT   coli dimethyl sulfoxide reductase by site-directed mutagenesis.";
RL   Biochemistry 30:8296-8305(1991).
CC   -!- FUNCTION: Electron transfer subunit of the terminal reductase
CC       during anaerobic growth on various sulfoxide and N-oxide
CC       compounds.
CC   -!- COFACTOR: BINDS 4 4FE-4S CLUSTERS.
CC   -!- SUBUNIT: THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA, THE
CC       REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSC, A
CC       MEMBRANE ANCHOR PROTEIN.
CC   -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC       'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J03412; AAA83844.1; -.
DR   EMBL; AE000191; AAC73981.1; -.
DR   EMBL; D90727; BAA35627.1; -.
DR   PIR; S03786; S03786.
DR   HSSP; P00195; 1CLF.
DR   EcoGene; EG10233; dmsB.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   Pfam; PF00037; fer4; 1.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; 1.
KW   Electron transport; 4Fe-4S; Iron-sulfur; Complete proteome.
FT   INIT_MET      0      0
FT   METAL        13     13       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        16     16       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        19     19       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        23     23       IRON-SULFUR 1 (4FE-4S) (BY SIMILARITY).
FT   METAL        66     66       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        69     69       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        74     74       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        78     78       IRON-SULFUR 2 (4FE-4S) (BY SIMILARITY).
FT   METAL        98     98       IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT   METAL       101    101       IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT   METAL       104    104       IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT   METAL       108    108       IRON-SULFUR 3 (4FE-4S) (BY SIMILARITY).
FT   METAL       125    125       IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT   METAL       128    128       IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT   METAL       140    140       IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT   METAL       144    144       IRON-SULFUR 4 (4FE-4S) (BY SIMILARITY).
FT   MUTAGEN     101    101       C->F,S,W,Y: LOSS OF ELECTRON TRANSFER
FT                                FROM MENAQUINOL TO DMSO.
FT   CONFLICT    169    169       P -> PRA (IN REF. 1).
```

```
SQ   SEQUENCE   204 AA;  22738 MW;  90425D307DEF2130 CRC64;

  Query Match             6.3%;  Score 83.5;  DB 1;  Length 204;
  Best Local Similarity   23.3%;  Pred. No. 4.6;
  Matches   45;  Conservative   27;  Mismatches   66;  Indels   55;  Gaps   10;

Qy         15 GVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIIISSVVPPIMFALE 74
              ||:|     |:  :  | :  ||       :    || | :: || ::           |
Db         51 GVWHQNVFAYY--LSISCNHCEDP----ACTKVCPSGAMHKREDGFVVVD---------E 95

Qy         75 RMC--TKYFHIEPQIVGPGMKTGLNIKYDNPK------EVGADRIVNAVAAIHLYGNPLI 126
               :|   :| |:      |        :|:  |      :   ||:      | :   ||
Db         96 DVCIGCRYCHMACPYGAP--------QYNETKGHMTKCDGCYDRVAEGKKPICVESCPLR 147

Qy        127 VVDFGTATTYCYIDENKQYMG--GAIAPGITISTEALYSRAAKLPRIEITRPDNIIGKNT 184
               :|||       ||| ::  |   |:||               |||   |:| ||: |
Db        148 ALDFGP------IDELRKKHGDLAAVAP--------------LPRAHFTKP-NIVIKPN 185

Qy        185 VSAMQSGILFGYV 197
               ::  :|   ||:
Db        186 ANSRPTGDTTGYL 198


RESULT 11
YHCI_HAEIN
ID   YHCI_HAEIN     STANDARD;      PRT;   300 AA.
AC   P44541;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Protein HI0144.
GN   HI0144.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION BY MASS SPECTROMETRY.
RX   MEDLINE=20137488; PubMed=10675023;
RA   Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA   Gray C., Fountoulakis M.;
RT   "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL   Electrophoresis 21:411-429(2000).
CC   -!- SIMILARITY: BELONGS TO THE ROK (NAGC/XYLR) FAMILY. STRONG, TO
CC       E.COLI YHCI.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32700; AAC21816.1; -.
DR   TIGR; HI0144; -.
DR   InterPro; IPR000600; ROK_family.
DR   Pfam; PF00480; ROK; 1.
DR   PROSITE; PS01125; ROK; 1.
KW   Complete proteome.
SQ   SEQUENCE   300 AA;  31918 MW;  8EDCF138D6EDC533 CRC64;

  Query Match             6.2%;  Score 83;  DB 1;  Length 300;
  Best Local Similarity   23.1%;  Pred. No. 8;
  Matches   37;  Conservative   28;  Mismatches   59;  Indels   36;  Gaps    8;

Qy          3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMIL-RSLFDHSGLMFEQIDGII 61
              | :|:| |      :  :|::|   :| | |     :       | : | |: |     | | :
Db          4 LALDIGGTKIAAAIVKNGEIEQRQQIHTPRENVVEGMHQALGKLLADYEG----QFDYVA 59

Qy         62 ISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVG--ADRIVNAVAAIH 119
              ::|                          |:  |: : |     ||| :|  |:  : |  | |
Db         60 VAST-------------------GIINNGILSAL-----NPKNLGGLAEFPLKASIAKH 94

Qy        120 LYGNPLIVVDFGTATTYC-YIDENKQYMGGAIAPGITIST 158
                 |: :::   | ||  |  :| : :   :   ||:||
Db         95 T-DKPIGLLNDAQAATYAEYQLQNSEQVSNFVF--ITVST 131


RESULT 12
GSA_CLOPE
ID   GSA_CLOPE     STANDARD;      PRT;   425 AA.
AC   Q9ZNC8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)
DE   (Glutamate-1-semialdehyde aminotransferase) (GSA-AT).
GN   HEML OR CPE1432.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC   Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 8237;
RX   MEDLINE=20049836; PubMed=10585141;
RA   Koyama M., Katayama S., Kaji M., Taniguchi Y., Matsushita O.,
RA   Minami J., Morita S., Okabe A.;
RT   "A Clostridium perfringens hem gene cluster contains a cysG(B)
RT   homologue that is involved in cobalamin biosynthesis.";
RL   Microbiol. Immunol. 43:947-957(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / Type A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
CC   -!- CATALYTIC ACTIVITY: (S)-4-amino-5-oxopentanoate = 5-
CC       aminolevulinate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; second step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB017186; BAA74784.1; -.
DR   EMBL; AP003190; BAB81138.1; ALT_INIT.
```

```
DR   HSSP; P24630; 2GSA.
DR   InterPro; IPR000954; Aminotran_3.
DR   InterPro; IPR004639; HemL.
DR   Pfam; PF00202; aminotran_3; 1.
DR   TIGRFAMs; TIGR00713; hemL; 1.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW   Porphyrin biosynthesis; Isomerase; Pyridoxal phosphate;
KW   Complete proteome.
FT   BINDING     265    265       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT   CONFLICT     73     73       E -> K (IN REF. 1).
FT   CONFLICT    323    323       N -> D (IN REF. 1).
FT   CONFLICT    356    356       I -> M (IN REF. 1).
FT   CONFLICT    422    422       T -> S (IN REF. 1).
SQ   SEQUENCE   425 AA;  46881 MW;  B23DFD62CE39B9E2 CRC64;

  Query Match             6.2%;  Score 83;  DB 1;  Length 425;
  Best Local Similarity   22.6%;  Pred. No. 12;
  Matches   43;  Conservative   21;  Mismatches   52;  Indels   74;  Gaps   10;

Qy         11 NTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIIISSVV---- 66
              ||::|:|:| |                     : ||:  |     | ||||  |
Db        175 NTLIGIYNDEKQ-------------------VEELFEKYG---NDIAGIIIEPVAGNMG 211

Qy         67 -----PPIMFALERMCTKY---------------------FHIEP------QIVGPGM 92
                   |  |  |  :| ||                       | ::|      :|:| |:
Db        212 VVKCDPKFMRKLRELCDKYGALLIFDEVMCGFRVAYKGAQTLFDVKPDLVTYAKIMGGGL 271

Qy         93 KTGLNIKYDNPKEV--GADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQY----- 145
                |    |   :|:     :    |  : ||| ||:  | ||    : ||  |
Db        272 PCG---AYGGRREIMENLSPLGGVYQAGTMSGNP-IVMSAGLATVK-KLYENPSYYNHIE 326

Qy        146 -MGGAIAPGI 154
               :|  :  |:
Db        327 KIGSKLEKGV 336


RESULT 13
CH60_TRIVA
ID   CH60_TRIVA     STANDARD;      PRT;   470 AA.
AC   Q95058;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Hydrogenosomal chaperonin HSP60 (Protein Cpn60) (GroEL protein) (Heat
DE   shock protein 60).
GN   HSP60.
OS   Trichomonas vaginalis.
OC   Eukaryota; Parabasalidea; Trichomonadida; Trichomonadidae;
OC   Trichomonadinae; Trichomonas.
OX   NCBI_TaxID=5722;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96382521; PubMed=8790385;
RA   Bui E.T., Bradley P.J., Johnson P.J.;
RT   "A common evolutionary origin for mitochondria and hydrogenosomes.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:9651-9656(1996).
CC   -!- FUNCTION: IMPLICATED IN HYDROGENOSOMAL PROTEIN IMPORT AND
CC       MACROMOLECULAR ASSEMBLY. MAY FACILITATE THE CORRECT FOLDING OF
CC       IMPORTED PROTEINS. MAY ALSO PREVENT MISFOLDING AND PROMOTE THE
CC       REFOLDING AND PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED
CC       UNDER STRESS CONDITIONS IN THE HYDROGENOSOME (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: HYDROGENOSOMAL.
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U26966; AAB17250.1; -.
DR   HSSP; P06139; 1GRL.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding; Hydrogenosome.
SQ   SEQUENCE   470 AA;  50614 MW;  6D501801045605AF CRC64;

  Query Match             6.2%;  Score 83;  DB 1;  Length 470;
  Best Local Similarity   18.8%;  Pred. No. 13;
  Matches   51;  Conservative   49;  Mismatches   76;  Indels   96;  Gaps   14;

Qy         28 IETSRH------KTED-EFGMILRSLFDHSGLMFEQIDGIIISSVVPP---IMFALERMC 77
              || ::|      |  | :||   |   ||  |  |::   ::|:: |    :|  |
Db          4 IEAAKHFTRAFAKARDLKFGSDAR---DHLLLGVEKLADAVVSTLGPKGRNVMIELP--- 57

Qy         78 TKYFHIEPQIVGPGMKTGLNIKY-DNPKEVGADRIVNAVAAIH----------------L 120
                   :  |::   |:    :|:: |  : :||  ::|     :                |
Db         58 ----YGPPKVTKDGVTVAKSIEFKDKWQNLGAQLVINVAQKTNDVAGDGTTTATLLTREL 113

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              |   :  |       :| ||      :  |:|:  :|:     |  : ::: || |
Db        114 YRESIKALSAG-------LDPNK------VRKGMTLRVDAVVKELEKSTK-KVSSPDEIF 159

Qy        181 GKNTVSAMQS----------------GILFGYVGQ--------VEGI------------ 203
                |:||  |                |::   :|:        |:|:
Db        160 NVATISANGSEKIGHLIADAFKAVGNEGVITVAMGKKFDHELETVQGMKIDRGYISSYFQ 219

Qy        204 --VKRMKWQAKQ------DLKVIATGGLAPLI 227
                | || : :       |:|: :   :||::
Db        220 NDTKSMKCEYENPYILITDIKINSFAQIAPIL 251


RESULT 14
GALE_METJA
ID   GALE_METJA     STANDARD;      PRT;   305 AA.
AC   Q57664;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Putative UDP-glucose 4-epimerase (EC 5.1.3.2) (Galactowaldenase) (UDP-
DE   galactose 4-epimerase).
GN   MJ0211.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8688087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT   jannaschii.";
RL   Science 273:1058-1073(1996).
CC   -!- CATALYTIC ACTIVITY: UDP-glucose = UDP-galactose.
CC   -!- COFACTOR: NAD.
CC   -!- PATHWAY: Galactose metabolism; third step.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR EPIMERASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U67477; AAB98196.1; -.
DR   HSSP; P09147; 1XEL.
DR   TIGR; MJ0211; -.
DR   InterPro; IPR001509; Epimerase_Dh.
DR   Pfam; PF01370; Epimerase; 1.
KW   Hypothetical protein; Isomerase; NAD; Galactose metabolism;
KW   Complete proteome.
FT   NP_BIND       1     32       NAD (POTENTIAL).
SQ   SEQUENCE   305 AA;  34542 MW;  2756773CF95D50BD CRC64;

  Query Match             6.1%;  Score 82;  DB 1;  Length 305;
  Best Local Similarity   24.7%;  Pred. No. 9.8;
  Matches   49;  Conservative   25;  Mismatches   52;  Indels   72;  Gaps   12;

Qy         99 KYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDEN----------KQYMG 147
              |||       |:|| | :   :|| |          |  :|||            :|:|
Db        104 KYD------IDKIVFASSGGAVYGEP----------NYLPVDENHPINPLSPYGLSKYVG 147

Qy        148 GAIAPGITISTEALYSRAAKLPRIE--ITRPDNIIGK--------------NTVSAMQS 190
                    ||:|   |  ||  | |  |: |:             : :   ||
Db        148 EEYI--------KLYNR---LYGIEYAILRYSNVYGERQDPKGEAGVISIFIDKMLKNQS 196

Qy        191 GILFG---------YVGQV-EGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
               |:||         ||| | :  :  : |  | ::  | ||    :    ::  ||:
Db        197 PIIFGDGNQTRDFVYVGDVAKANLMALNW--KNEIVNIGTGKETSV----NELFDIIKHE 250

Qy        241 LTLKGLELIYERNRVGSV 258
              :  :| | ||:: | | |
Db        251 IGFRG-EAIYDKPREGEV 267


RESULT 15
POLG_HCVB
ID   POLG_HCVB      STANDARD;      PRT;  3898 AA.
AC   P21530;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Genome polyprotein.
OS   Hog cholera virus (strain Brescia) (Swine fever virus).
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Pestivirus.
OX   NCBI_TaxID=11098;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90281581; PubMed=2162104;
RA   Moormann R.J.M., Warmerdam P.A.M., van der Meer B., Schaaper W.M.M.,
RA   Wensvoort G., Hulst M.M.;
RT   "Molecular cloning and nucleotide sequence of hog cholera virus
RT   strain Brescia and mapping of the genomic region encoding envelope
RT   protein E1.";
RL   Virology 177:184-198(1990).
CC   -!- FUNCTION: PESTIVIRUS P80 (P125) MAY BE A BIFUNCTIONAL PROTEIN
CC       WITH HELICASE AND PROTEASE ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: THE GP51-GP54 PROTEIN IS ANCHORED TO THE
CC       VIRAL ENVELOPE.
CC   -!- SIMILARITY: TO BOVINE VIRAL DIARRHEA VIRUS GENOME POLYPROTEIN.
CC   -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S31.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M31768; AAA43843.1; -.
DR   PIR; A35317; GNWVHB.
DR   HSSP; P27958; 1A1V.
DR   MEROPS; C53.001; -.
DR   MEROPS; S31.001; -.
DR   InterPro; IPR000280; CDvir_endptseP80.
DR   InterPro; IPR001410; DEAD.
DR   InterPro; IPR002166; HCV_RdRP.
DR   InterPro; IPR001650; Helicase_C.
DR   InterPro; IPR001568; RNase_T2.
DR   Pfam; PF00271; helicase_C; 1.
DR   Pfam; PF00998; HCV_RdRP; 1.
DR   PRINTS; PR00729; CDVENDOPTASE.
DR   SMART; SM00490; HELICc; 1.
DR   PROSITE; PS00531; RNASE_T2_2; UNKNOWN_1.
KW   Polyprotein; Glycoprotein; Transmembrane; Hydrolase; Serine protease;
KW   Helicase.
FT   CHAIN        ?1    ?267       CAPSID PROTEIN C (POTENTIAL).
FT   CHAIN      ?268    ?500       GP42 (E2) (POTENTIAL).
FT   CHAIN      ?501    ?689       GP31 (E3) (POTENTIAL).
FT   CHAIN      ?690   ?1060       GP51-GP54 (ENVELOPE PROTEIN E1).
FT   CHAIN     ?1611   ?2111       P80 (POTENTIAL).
FT   TRANSMEM   1032   1048       POTENTIAL.
FT   ACT_SITE   1658   1658       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1695   1695       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   ACT_SITE   1752   1752       CHARGE RELAY SYSTEM (BY SIMILARITY).
FT   CARBOHYD    157    157       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    269    269       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    278    278       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    332    332       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    362    362       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    410    410       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    425    425       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    500    500       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    805    805       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    810    810       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    874    874       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    918    918       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    949    949       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1713   1713       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2134   2134       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2217   2217       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2419   2419       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2494   2494       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2787   2787       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2815   2815       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2891   2891       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3103   3103       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3211   3211       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3316   3316       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3689   3689       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3698   3698       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3794   3794       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   3898 AA;  438423 MW;  EC6EB207A09D59FD CRC64;

  Query Match             6.1%;  Score 82;  DB 1;  Length 3898;
  Best Local Similarity   21.2%;  Pred. No. 1.9e+02;
  Matches   46;  Conservative   34;  Mismatches   73;  Indels   64;  Gaps   11;

Qy         20 GKLEYHW-RIETSRHKTEDEFGMILR--------SLFDHSGLMFEQID---GIIISSVVP 67
              |  :||:  ::  |:  ||   :           ||::   ||  |::    ::||  :|
Db       2113 GSKDYHYDLLQAQRYGIEDGINITKSFREMNYDWSLYEEDSLMITQLEILNNLLISEELP 2172

Qy         68 PIMFALERMCTKYFHIEP-----------QIVGPGMKTG-LNIKYDNPKEVGADRIVNA 114
                |:: :  :  | ||             :: | :| | :   |||   : | :: :
Db       2173 ---MAVKNIMARTDHPEPIQLAYNSYETQVPVLFPKIKNGEVTDSYDNYTFLNARKLGDD 2229

Qy        115 VAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIA-----PGITISTEALYSRAAKLP 169
              |                  | |  |::     :      ||   : ||   || |
Db       2230 V-----------------PPYVYATEDEDLAVELLGLDWPDPGNQGTVEA--GRALK-- 2267

Qy        170 RIEITRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKR 206
                    ::| :|        ||||||  : : ||
Db       2268 --------QVVGLSTAENALLVALFGYVG-YQALSKR 2295
```

Search completed: June 24, 2003, 21:51:02
Job time : 6.48734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.41033 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1114	83.4	233	2	S66100	conserved hypothet
2	987	73.9	254	2	F83660	hypothetical prote
3	852	63.8	259	2	AF1102	conserved hypothet
4	847	63.4	259	2	AF1464	conserved hypothet
5	775	58.1	273	2	E97293	probable transcrip
6	655	49.1	265	2	T36391	hypothetical prote
7	577	43.2	261	2	B87489	transcription acti
8	519.5	38.9	274	2	H86937	conserved hypothet
9	494.5	37.0	272	2	A70955	hypothetical prote
10	428	32.1	262	2	E75516	conserved hypothet
11	405.5	30.4	246	2	D72320	conserved hypothet
12	327.5	24.5	273	2	D71326	conserved hypothet
13	271.5	20.3	262	2	F70165	conserved hypothet
14	170	12.7	276	2	AI2292	hypothetical prote
15	163	12.2	257	2	S75559	hypothetical prote
16	154.5	11.6	229	2	E70465	hypothetical prote
17	150	11.2	592	2	B81009	BirA protein/Bvg a
18	150	11.2	592	2	H82031	probable biotin-[a
19	134.5	10.1	242	2	A82637	conserved hypothet
20	133	10.0	248	2	H83111	hypothetical prote
21	129.5	9.7	267	2	I40327	baf protein - Bord
22	117	8.8	224	2	A99571	conserved hypothet
23	116.5	8.7	223	2	G71887	hypothetical prote
24	109.5	8.2	223	2	F64627	hypothetical prote
25	109	8.2	209	2	H81382	hypothetical prote
26	104.5	7.8	597	2	B69251	probable electron
27	95	7.1	467	2	AE3142	hypothetical prote
28	95	7.1	520	2	G98145	probable aminotran
29	93	7.0	1014	2	H71602	protein with DnaJ
30	91.5	6.9	828	2	G82583	bifunctional aspar
31	90.5	6.8	4452	1	YGBSG2	gramicidin S synth
32	88	6.6	256	2	H83002	imidazoleglycerol-
33	87.5	6.6	287	2	D90540	glucokinase (gluco
34	87	6.5	336	2	AE0625	dihydroorotate deh
35	86.5	6.5	205	2	AC0612	anaerobic dimethyl
36	86.5	6.5	205	2	AG0680	probable dimethyl
37	86.5	6.5	243	2	D64300	(R)-2-hydroxygluta
38	86	6.4	336	1	S13824	dihydroorotate oxi
39	85	6.4	336	1	DEECDO	dihydroorotate oxi
40	85	6.4	336	2	E90757	dihydro-orotate de
41	85	6.4	336	2	C85621	dihydro-orotate de
42	84.5	6.3	317	2	E96792	probable cinnamoyl
43	84	6.3	443	1	JC5298	protein kinase (EC
44	83.5	6.3	205	2	B85615	hypothetical prote
45	83.5	6.3	205	2	D90751	hypothetical prote

ALIGNMENTS

RESULT 1
S66100
conserved hypothetical protein yacB - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S66100; E69740
R;Ogasawara, N.; Nakai, S.; Yoshikawa, H.
DNA Res. 1, 1-14, 1994
A;Title: Systematic sequencing of the 180 kilobase region of the Bacillus subtilis
A;Reference number: S65967; MUID:96051385; PMID:7584024
A;Accession: S66100
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <OGA>
A;Cross-references: EMBL:D26185; NID:g467326; PIDN:BAA05305.1; PID:d1005847; PID:g4
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 19
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardi
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scar
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchi
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshi
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subti
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69740
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-233 <KUN>
A;Cross-references: GB:Z99104; GB:AL009126; NID:g2632267; PIDN:CAB11846.1; PID:e118
A;Experimental source: strain 168
C;Genetics:
A;Gene: yacB
A;Start codon: TTG
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 83.4%; Score 1114; DB 2; Length 233;
Best Local Similarity 99.5%; Pred. No. 3.6e-88;
Matches 213; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy         121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy         181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214
                |||||||||||||||||||||||||||||||||
Db         181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214


RESULT 2
F83660
hypothetical protein BH0086 [imported] - Bacillus halodurans (strain C-125)
C;Species: Bacillus halodurans
C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C;Accession: F83660
R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
  Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
  Reference number: A83650; MUID:20512582; PMID:11058132
A;Accession: F83660
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-254 <STO>
A;Cross-references: GB:AP001507; GB:BA000004; NID:g10172612; PIDN:BAB03805.1; GSPDB:GN0
A;Experimental source: strain C-125
C;Genetics:
A;Gene: BH0086
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             73.9%;  Score 987;  DB 2;  Length 254;
  Best Local Similarity   72.3%;  Pred. No. 3.2e-77;
  Matches  183;  Conservative   33;  Mismatches   37;  Indels    0;  Gaps    0;

Qy           1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
                 ::||||||||||||||| |  | :|||: ||| ||||| | :|||||:|| |: ||||
Db           1 MILVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGI 60

Qy          61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
                 :|||||||:||:||:|| |||| | |:|||:|||||||||||||||||||||||| |
Db          61 VISSVVPPMMFSLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRIVNAVAAIEL 120

Qy         121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
                 || | |||||||||||| |:| ||| || ||||| ||||||| ||:||||||| :|  ::
Db         121 YGYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQVV 180

Qy         181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
                 | ||: :||||| :||| ||:|:||||| ||: : |||||||| ||  ||: ||::| |
Db         181 GTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240

Qy         241 LTLKGLELIYERN 253
                ||||||:|||::|
Db         241 LTLKGLQLIYKKN 253


RESULT 3
AF1102
conserved hypothetical protein lmo0221 [imported] - Listeria monocytogenes (strain EGD-e
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1102
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1102
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAD00748.1; PID:g16409586; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo0221
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             63.8%;  Score 852;  DB 2;  Length 259;
  Best Local Similarity   63.4%;  Pred. No. 1.2e-65;
  Matches  161;  Conservative   37;  Mismatches   56;  Indels    0;  Gaps    0;

Qy           1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
                 ::|||||||||  :|||   ||  |||: | ||:| || || : : | :: |    | ||
Db           1 MILVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy          61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
                 |||||||||| |:| || :||:| | |||||:|||||:| |||:|:|:|||||||||
Db          61 IISSVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy         121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
                 || |:||||||||||:|||||:  | |||||||| |||||||:|||||||::|     ||
Db         121 YGTPVIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQII 180

Qy         181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
                 ||:|||:||:|| :|:||| |||:  || |:     |:|||||| :|  :|  :||:|||
Db         181 GKSTVSSMQAGIFYGFVGQCEGIIAEMKKQSNASPVVVATGGLARMITEKSSAVDILDPF 240

Qy         241 LTLKGLELIYERNR 254
                ||||||||:| ||:
Db         241 LTLKGLELLYRRNK 254


RESULT 4
AF1464
conserved hypothetical protein lin0253 [imported] - Listeria innocua (strain Clip112
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AF1464
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1464
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-259 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95486.1; PID:g16412682; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0253
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             63.4%;  Score 847;  DB 2;  Length 259;
  Best Local Similarity   62.6%;  Pred. No. 3.3e-65;
  Matches  159;  Conservative   39;  Mismatches   56;  Indels    0;  Gaps    0;

Qy           1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
                 ::|||||||||  :||| : ||  |||: | ||:| || || : : | :: |    | ||
Db           1 MILVIDVGNTNCTVGVYKEQKLLRHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy          61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
                 |||||||||| |:| || :||:| | |||||:|||||:| |||:|:|:|||||||||
Db          61 IISSVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy         121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
                 || |:||||||||||:|||||   | |||||||| |||||||:|||||||::|     ||
Db         121 YGTPVIVVDFGTATTFCYIDEAGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQII 180
```

```
Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||:||::||::|| :|::|| |||:  || |:     |:|||||| :|  :|  :||:|||
Db        181 GKSTVASMQAGIFYGFIGQCEGIIAEMKKQSNTSPVVVATGGLARMITEKSSAVDILDPF 240

Qy        241 LTLKGLELIYERNR 254
              ||||||||:| ||:
Db        241 LTLKGLELLYRRNK 254


RESULT 5
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             58.1%;  Score 775;  DB 2;  Length 273;
  Best Local Similarity   56.4%;  Pred. No. 5.4e-59;
  Matches  146;  Conservative   55;  Mismatches   54;  Indels    4;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::||:|||||| |||:|:| ||   ||: |   :: ||:|: : :||    |    ::|:
Db         12 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGV 71

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| ||::|| |  ||| | | :||||:|||:|||||||||||||||||||||  :
Db         72 IISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVAAHEI 131

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              |   ||::|||||||:| : ||  |:|||| ||| :|:|||: :||||||:|: :|   |
Db        132 YKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKPAYAI 191

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLK----VIATGGLAPLIANESDCIDI 236
               |||:|::|||||::||:|||  ||:||| : ::: :    |:|||||| ||: |:  :|:
Db        192 CKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVATGGLAKLISEEAKNVDV 251

Qy        237 VDPFLTLKGLELIYERNRV 255
              ::|||||:|| :|||:|||
Db        252 INPFLTLEGLRIIYEKNRV 270


RESULT 6
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE94.31c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             49.1%;  Score 655;  DB 2;  Length 265;
  Best Local Similarity   51.0%;  Pred. No. 1e-48;
  Matches  133;  Conservative   46;  Mismatches   74;  Indels    8;  Gaps    3;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ---- 56
              :|| ||||||:||||::    :  |||| |   :| ||  ::|: |     |: ::
Db          1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDG 60

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              |||| | : || ::  |  :  :|:   | : | ||:|||: |  |:||||||||:|||
Db         61 IDGIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPILTDHPKEVGADRIINAV 120

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||: ||| | ||||||||||:  :    :|:|| ||||| || |||  : |:| :||: |
Db        121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVAR 180

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQA---KQDLKVIATGGLAPLIANESD 232
              | ::|||||| ||||||::|: |||:|:| ||  :      |: |||||||||::  ||
Db        181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGGLAPMVLGESS 240

Qy        233 CIDIVDPFLTLKGLELIYERN 253
              ||  :|:||| || |:||||
Db        241 VIDEHEPWLTLMGLRLVYERN 261


RESULT 7
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelber
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.;
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser,
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             43.2%;  Score 577;  DB 2;  Length 261;
  Best Local Similarity   44.4%;  Pred. No. 5e-42;
  Matches  114;  Conservative   50;  Mismatches   91;  Indels    2;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|| |: |||||: .::         ||  |   :| ||: : |  |    || |  || :
Db          2 MLLAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAV 61

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||  :| |  :  :||::|| ::|    | |:::: : | | ||||:|||: |  :
Db         62 IISSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 121

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRP--DN 178
              |  ||:|:| |||||:  :  :  : || ||||| :| :||:  |||||||  | ||  :
Db        122 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 181

Qy        179 IIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVD 238
              |:| :|||||||: :||:  :||:| |:| :  : : ||||||:| |    :| ||  |
Db        182 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFD 241

Qy        239 PFLTLKGLELIYERNRV 255
              ||::||  || || :
Db        242 SDLTIRGLLEIYRRNTI 258


RESULT 8
```

H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 38.9%; Score 519.5; DB 2; Length 274;
Best Local Similarity 41.0%; Pred. No. 4.5e-37;
Matches 109; Conservative 53; Mismatches 89; Indels 15; Gaps 5;

```
Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  :|:  |        |:
Db          1 MLLAIDVRNTHTVVGLLSGSKEHAKVVQQWRIRTESEVTADELALIIDGLIGDDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : : ||::||: :  ||||||||||||| :
Db         58 LAGAAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              || | :|   ||||||::     :    :::|||||||: :|::|  :|:| | |:|: |
Db        118 AAFHKFGQAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQ---DL----KVIATGGLAPLIA 228
              | :::|||||  ||:|:::||: | |:|:| ||:   ::   ||     |:|||  |||:
Db        178 PRSVVGKNTVECMQAGVVFGFAGLVDGLVGRMRQDVEEFSGDLGNRVAVVATGHTAPLLL 237

Qy        229 NESDCIDIVDPFLTLKGLELIYERNR 254
              |   :|  |  ||| || |::||||
Db        238 PELHTVDHYDRHLTLHGLRLVFERNR 263
```

RESULT 9
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
;Species: Mycobacterium tuberculosis
;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70955
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70955
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-272 <COL>
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113976
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3600c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 37.0%; Score 494.5; DB 2; Length 272;
Best Local Similarity 38.8%; Pred. No. 6.3e-35;
Matches 102; Conservative 55; Mismatches 93; Indels 13; Gaps 4;

```
Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  : :  |        |:
Db          1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : : ||::||: :  ||||||||||||| :
Db         58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||   :    ||||||::     :    :::|||||||: :|::|  :|:| | |:|: |
Db        118 AAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
              | :::|||||  ||:| :||: | |:|:| |::     :    |: ::|||  |||:  |
Db        178 PRSVVGKNTVECMQAGAVFGFAGLVDGLVGRIREDVSGFSVDHDVAIVATGHTAPLLLPE 237

Qy        231 SDCIDIVDPFLTLKGLELIYERN 253
              :|  |  |||:|| |::|||
Db        238 LHTVDHYDQHLTLQGLRLVFERN 260
```

RESULT 10
E75516
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
C;Accession: E75516
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: E75516
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-262 <WHI>
A;Cross-references: GB:AE001905; GB:AE000513; NID:g6458144; PIDN:AAF10040.1; PID:g64
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR0461
A;Map position: 1
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

Query Match 32.1%; Score 428; DB 2; Length 262;
Best Local Similarity 37.4%; Pred. No. 3e-29;
Matches 98; Conservative 51; Mismatches 99; Indels 14; Gaps 6;

```
Qy          2 LLVIDVGNTNTVLGVYH-DGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              || :|:||| ||||:    | | : ||| |:|    |:  : |  ||  :|    :
Db          6 LLAVDIGNTTTVLGLADASGALTHTWRIRTNREMLPDDLALQLHGLFTLAGAPIPR--AA 63

Qy         61 IISSVVPPI----MFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVA 116
              ::||| ||:      ||:|    :| |:   |       : :: | |  |||||: |
Db         64 VLSSVAPPVGENYALALKR----HFMIDAFAVSAENLPDVTVELDTPGSVGADRLCNLFG 119

Qy        117 AIHLYG--NPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEIT 174
              |    |  :  :||||||:| :  :   ::::|| :| |  :| :||::|||||||| :
Db        120 AEKYLGGLDYAVVVDFGTSTNFDVVGRGRRFLGGILATGAQVSADALFARAAKLPRITLQ 179

Qy        175 RPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCI 234
              |:  ||||||  |:|||::|||   |:|:::|:: :   :   :|||| :  :      |
Db        180 APETAIGKNTVHALQSGLVFGYAEMVDGLLRRIRAELPGEAVAVATGGFSRTVQGICQEI 239

Qy        235 DIVDPFLTLKGL-ELIYERNRV 255
              |  |  |||:|| ||   |: |
Db        240 DYYDETLTLRGLVELWASRSEV 261
```

RESULT 11
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima

```
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498141
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             30.4%;  Score 405.5;  DB 2;  Length 246;
  Best Local Similarity   39.2%;  Pred. No. 2.4e-27;
  Matches  100;  Conservative   42;  Mismatches   88;  Indels   25;  Gaps    9;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDH----SGLMFEQ 56
              : |::|||||::|  :  |||    ||: |   :|||       || |     |    :
Db          1 MYLLVDVGNTHSVFSITEDGKTFRRWRLSTGVFQTEDE-------LFSHLHPLLGDAMRE 53

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGL---NIKYDNPKEVGADRIVN 113
              | || ::|||     :||   ||||| |  :    | |    |:|  || ||||||: |
Db         54 IKGIGVASVVPTQNTVIERFSQKYFHISP--IWVKAKNGCVKWNVK--NPSEVGADRVAN 109

Qy        114 AVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEI 173
               || : ||   |::| |||||   :  |  | |||| ||  :   :|:   |||| :|:
Db        110 VVAFVKEYGKNGIIIDMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGTAKLPLVEV 168

Qy        174 TRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233
                | ::||:|   :: |:: | |  :|||: |:| :   || |: ||| : ::    |
Db        169 KPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIK-EVYGDLPVVLTGGQSKIV---KDM 224

Qy        234 I--DIVDPFLTLKGL 246
              |  :| |  ||:||:
Db        225 IKHEIFDEDLTIKGV 239


RESULT 12
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C;Accession: D71326
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71326
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <COL>
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0431
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             24.5%;  Score 327.5;  DB 2;  Length 273;
  Best Local Similarity   29.9%;  Pred. No. 1.3e-20;
  Matches   75;  Conservative   64;  Mismatches  105;  Indels    7;  Gaps    4;

Qy          2 LLVIDVGNTNTVLGVY--HDGK--LEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQI 57
              :|:|||||:: | |:   : |:  :   :|:     ||:||: ::: :| : :|:    :
Db          1 MLLIDVGNSHVVFGIQGENGGRVCVRELFRLAPDARKTQDEYSLLIHALCERAGVGRASL 60

Qy         58 DGIIISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNP--KEVGADRIVNAV 115
                |||||| :   :     :  ::| : ||     | ::   |   |:| | : |||
Db         61 RDAFISSVVPVLTKTIADAVAQISGVQPVVFGPWAYEHLPVRIPEPVRAEIGTDLVANAV 120

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||   : :  :||| ||| |:  :|      | |||||:  : ::|::  |:|| : :
Db        121 AAYVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPLVPLAL 180

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCID 235
              ||:::||:|  |:|:|:: | :  :  :: : : :       : ||||: | ::| |
Db        181 PDSVLGKDTTHAVQAGVVRGTLFVIRAMIAQCQKELGCRCAAVITGGLSRLFSSEVD-FP 239

Qy        236 IVDPFLTLKGL 246
               :|  ||| ||
Db        240 PIDAQLTLSGL 250


RESULT 13
F70165
conserved hypothetical protein BB0527 - Lyme disease spirochete
C;Species: Borrelia burgdorferi (Lyme disease spirochete)
C;Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C;Accession: F70165
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.;
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.;
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch,
Nature 390, 580-586, 1997
A;Authors: Smith, H.O.; Venter, J.C.
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A;Reference number: A70100; MUID:98065943; PMID:9403685
A;Accession: F70165
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-262 <KLE>
A;Cross-references: GB:AE001154; GB:AE000783; NID:g2688431; PIDN:AAC66882.1; PID:g2
A;Experimental source: strain B31

  Query Match             20.3%;  Score 271.5;  DB 2;  Length 262;
  Best Local Similarity   27.3%;  Pred. No. 8.2e-16;
  Matches   70;  Conservative   57;  Mismatches  108;  Indels   21;  Gaps    5;

Qy          3 LVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGIII 62
              |:||:|||:    :: | ::    :::|:     ||        || :      :: : |
Db          9 LIIDIGNTSIAFALFKDNQVNLPIKMKTNLMLRYDEVYSFFEENFDFN------VNKVFI 62

Qy         63 SSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYD---NPKE-----VGADRIVNA 114
              ||||| :    : :   :| |.:|  :       | :: ||   ||  :    :|:|   |
Db         63 SSVVPILNETFKNVIFSFFKIKPLFI------GFDLNYDLTFNPYKSDKFLLGSDVFANL 116

Qy        115 VAAIHLYG-NPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEI 173
              ||||  |    ::||| ||| |   :       :|| |  |  |:  :|   |  : :  |
Db        117 VAAIENYSFENVLVVDLGTACTIFAVSRQDGILGGIINSGPLINFNSLLDNAYLIKKFPI 176

Qy        174 TRPDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDC 233
              : |:|:: : |  :: ||: : |   :||: : :|   |:   :| ||| | || :  :
Db        177 STPNNLLERTTSGSVNSGLFYQYKYLIEGVYRDIKQMYKKKFNLIITGGNADLILSLIEI 236

Qy        234 IDIVDPFLTLKGLELI 249
               | :  ||::|: ::
Db        237 EFIFNIHLTVEGVRIL 252


RESULT 14
AI2292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Taba
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacteriu
```

```
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

  Query Match            12.7%;  Score 170;  DB 2;  Length 276;
  Best Local Similarity   24.2%;  Pred. No. 4.5e-07;
  Matches   66;  Conservative   50;  Mismatches  115;  Indels   42;  Gaps   10;

Qy          3 LVIDVGNTNTVLGVYHDGKLEYHWRIE---------TSRHKTEDEFGMILRSLFDHSGLM 53
              | :::||:     ::    ||: |  |             :|: | |   : :|
Db         15 LALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGETKLEVGSEEKEIF----FT 70

Qy         54 FEQID------GIIISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVG 107
              |  :        : |:|||      |     | ::    :    :  ||  |     :|
Db         71 FFPLPPAPCPLPLFIASVVPQQTVLWE----NYLNVRVITLD---QIPLNNIYPT---LG 120

Qy        108 ADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAK 167
               || :    |   :| |::|:| ||| |:   |  |  :|||| ||: :   :|  :  :
Db        121 IDRALALWGAGMSWGFPVLVIDAGTALTFTAADGGKNLVGGAILPGVGLQFASLGQQTGQ 180

Qy        168 LPRIEITRPDNI---IGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAK-QDLKVIATGGL 223
              ||::|:    ::      ||  |:|||:::  :  :    :  :| :   | ||   ||
Db        181 LPQVEMEAIKSLPPRFALNTTEAIQSGVIYTLIAGMRDFTE--EWLSLFPDGKVAIKGGD 238

Qy        224 APLIANESDCID-------IVDPFLTLKGLELI 249
               |: |    :        ||:| |   |:: |
Db        239 RILLLNYLQALYPDLAARLIVEPNLIFWGMQTI 271


RESULT 15
S75559
hypothetical protein slr0812 - Synechocystis sp. (strain PCC 6803)
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C;Accession: S75559
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75559
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-257 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18120.1; PID:d101885
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
A;Start codon: GTG

  Query Match            12.2%;  Score 163;  DB 2;  Length 257;
  Best Local Similarity   28.0%;  Pred. No. 1.6e-06;
  Matches   58;  Conservative   42;  Mismatches   79;  Indels   28;  Gaps   10;

Qy         60 IIISSVVPPIMFALERMCTKYFHI-EPQIVGPGMKTGLNIKYDN-PKEVGADRIVNAVAA 117
              ::::|||||          |: : : :|:|:     |  |:   |     | || :  :
Db         64 LMLASVVPE--------QTEVWRVYQPKIL-----TLKNLPLVNLYPSFGIDRALAGLGT 110

Qy        118 IHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR-- 175
                 || | :||| ||| |    |::|: :|||| ||: :    |  | | ||::|: :
Db        111 GLTYGFPCLVVDGGTALTITGFDQDKKLVGGAILPGLGLQLATLGDRLAALPKLEMDQLT 170

Qy        176 --PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAK-QDLKVIATGG----LAPLIA 228
                 ||     :| ||: ||:::| :| ::  ::   ||       :: |||    |   :
Db        171 ELPDR-WALDTPSAIFSGVVYGVLGALQSYLQ--DWQKLFPGAAMVITGGDGKILHGFLK 227

Qy        229 NESDCIDIV-DPFLTLKGLELIYERNR 254
                 |  : :  |  |   |:  |:  :|
Db        228 EHSPNLSVAWDDNLIFLGMAAIHHGDR 254


Search completed: June 24, 2003, 22:02:56
Job time : 11.4103 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 26.4883 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-2
Perfect score: 1335
Sequence: 1 LLLVIDVGNTNTVLGVYHDG.....PFLTLKGLELIYERNRVGSV 258

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1335	100.0	258	22	AAU01243	B. subtilis novel
2	1335	100.0	258	23	AAU91149	Bacillus subtilis
3	1114	83.4	233	23	AAU91163	Pantothenate kinas
4	1046	78.4	258	23	AAU91172	Pantothenate kinas
5	1034	77.5	262	23	AAU91170	Pantothenate kinas
6	987	73.9	254	23	AAU91171	Pantothenate kinas
7	852	63.8	259	23	ABB47661	Listeria monocytog
8	794.5	59.5	256	23	AAU91175	Pantothenate kinas
9	756	56.6	255	23	AAU91154	Geobacter sulfurre
10	655	49.1	265	23	AAU91151	Streptomyces coeli
11	646.5	48.4	250	23	AAU91150	Clostridium acetob
12	577	43.2	260	23	AAU91173	Pantothenate kinas
13	537	40.2	258	23	AAU91153	Rhodobacter capsul
14	515	38.6	219	23	AAU91176	Pantothenate kinas
15	494.5	37.0	272	23	AAU91152	Mycobacterium tube
16	493.5	37.0	272	22	AAG81225	Mycobacterium tube
17	428	32.1	262	23	AAU91155	Deinococcus radiop
18	405.5	30.4	246	23	AAU91156	Thermotoga maritim
19	338.5	25.4	212	23	AAU91177	Pantothenate kinas
20	327.5	24.5	273	23	AAU91157	Treponema pallidum
21	319	23.9	257	23	AAU91174	Pantothenate kinas
22	271.5	20.3	262	23	AAU91158	Borrelia burgdorfe
23	207	15.5	244	23	AAU91168	Pantothenate kinas
24	203	15.2	241	23	AAU91179	Pantothenate kinas
25	163	12.2	249	23	AAU91182	Pantothenate kinas
26	163	12.2	257	23	AAU91160	Synechocystis pant
27	154.5	11.6	229	23	AAU91159	Aquifex aeolicus p
28	154	11.5	249	23	AAU91178	Pantothenate kinas
29	151	11.3	455	20	AAY38617	Neisseria gonorrho
30	151	11.3	455	21	AAY74908	Neisseria gonorrhe
31	151	11.3	460	23	AAU91167	Pantothenate kinas
32	151	11.3	592	20	AAY38618	Neisseria gonorrho
33	151	11.3	592	21	AAY74911	Neisseria gonorrhe
34	150	11.2	389	21	AAY74909	Neisseria meningit
35	150	11.2	455	21	AAY74910	Neisseria meningit
36	150	11.2	592	20	AAY38615	Neisseria meningit
37	150	11.2	592	20	AAY38616	Neisseria meningit
38	150	11.2	592	21	AAY74912	Neisseria meningit
39	150	11.2	592	21	AAY74913	Neisseria meningit
40	150	11.2	592	23	AAU91166	Pantothenate kinas
41	150	11.2	592	23	AAU91169	Pantothenate kinas
42	134.5	10.1	242	23	AAU91180	Pantothenate kinas
43	133	10.0	248	23	AAU91164	Pantothenate kinas
44	129.5	9.7	267	23	AAU91162	Bordella pertussis
45	114.5	8.6	189	20	AAY38614	Neisseria meningit

ALIGNMENTS

RESULT 1
AAU01243
ID AAU01243 standard; Protein; 258 AA.
XX
AC AAU01243;
XX
DT 18-JUL-2001 (first entry)
XX
DE B. subtilis novel pantothenate kinase encoded by the gene coaX.
XX
KW Pantothenate kinase; coaX; pantothenate biosynthesis; vitamin B5;
KW nutritional supplement; panto-compound; pantoate.
XX
OS Bacillus subtilis.
XX
PN WO200121772-A2.
XX
PD 29-MAR-2001.
XX
PF 21-SEP-2000; 2000WO-US25993.
XX
PR 21-SEP-1999; 99US-0400494.
PR 07-JUN-2000; 2000US-0210072.
PR 28-JUL-2000; 2000US-0221836.
PR 24-AUG-2000; 2000US-0227860.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS.
XX
PI Yocum RR, Patterson TA, Hermann T, Pero JG;
XX
DR WPI; 2001-218644/22.
DR N-PSDB; AAS00984.
XX

```
PT   New recombinant microorganism which overexpress a Bacillus subtilis
PT   pantothenate biosynthetic enzyme, useful for the high yield production
PT   of panto-compounds such as pantothenate and pantoate -
XX
PS   Example 14; Fig 23; 292pp; English.
XX
CC   The sequence represents a novel B. subtilis pantothenate kinase (encoded
CC   by gene coaX), an enzyme of the pantothenate biosynthetic pathway.
CC   Pantothenate, also known as vitamin B5, is used as a nutritional
CC   supplement in mammals and humans. The invention concerns methods of
CC   producing recombinant microorganisms overexpressing at least one Bacillus
CC   subtilis pantothenate biosynthetic enzyme. The microorganisms and methods
CC   of producing them are useful for producing a panto-compound such as
CC   pantothenate or pantoate, which is a nutritional requirement for
CC   livestock and humans. The methods are also useful for the identification
CC   of pantothenate kinase modulators. Panto-compounds are produced at a
CC   significantly higher yield than prior art methods and can be produced
CC   independent of the need to feed precursors which decreases expense.
XX
SQ   Sequence   258 AA;

  Query Match             100.0%;  Score 1335;  DB 22;  Length 258;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-142;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240

Qy        241 LTLKGLELIYERNRVGSV 258
              ||||||||||||||||||
Db        241 LTLKGLELIYERNRVGSV 258


RESULT 2
AAU91149
ID   AAU91149 standard; Protein; 258 AA.
XX
AC   AAU91149;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Bacillus subtilis pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus subtilis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54168.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 10; Page 67-68; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   258 AA;

  Query Match             100.0%;  Score 1335;  DB 23;  Length 258;
  Best Local Similarity   100.0%;  Pred. No. 8.2e-142;
  Matches  258;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240

Qy        241 LTLKGLELIYERNRVGSV 258
              ||||||||||||||||||
Db        241 LTLKGLELIYERNRVGSV 258


RESULT 3
AAU91163
ID   AAU91163 standard; Protein; 233 AA.
XX
AC   AAU91163;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #1.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus subtilis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
```

```
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54169.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Disclosure; Page 81-82; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   233 AA;

  Query Match             83.4%;  Score 1114;  DB 23;  Length 233;
  Best Local Similarity   99.5%;  Pred. No. 5.9e-117;
  Matches  213;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214
              |||||||||||||||||||||||||||||||||
Db        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQD 214


RESULT 4
AAU91172
ID   AAU91172 standard; Protein; 258 AA.
XX
AC   AAU91172;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #10.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus stearothermophilus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54193.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 10; Page 101-102; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   258 AA;

  Query Match             78.4%;  Score 1046;  DB 23;  Length 258;
  Best Local Similarity   78.3%;  Pred. No. 3.2e-109;
  Matches  198;  Conservative   32;  Mismatches   23;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :: |:|||||||||||||   :|::|||||||| |||||:||::::| :| || |  | ||
Db          1 MIFVLDVGNTNTVLGVYDGDELKHHWRIETSRSKTEDEYGMMIKALLNHVGLQFSDIRGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||||||||||| |||||:| |||||:||||:||||||:|||||||||||| |||
Db         61 IISSVVPPIMFALERMCLKYFHIKPLIVGPGIKTGLDIKYDNPREVGADRIVNAVAGIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||:|||:||||||||||||:|:|||||||||||| ||||||::|||||||||| |||:||
Db        121 YGSPLIIVDFGTATTYCYINEHKQYMGGAIAPGIMISTEALFARAAKLPRIEIARPDDII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              |||||||||:|||:|||||||||| ||| ::|   |||||||||||||:||| ||:||||
Db        181 GKNTVSAMQAGILYGYVGQVEGIVSRMKAKSKIPPKVIATGGLAPLIASESDIIDVVDPF 240

Qy        241 LTLKGLELIYERN 253
              ||| ||:|:||:|
Db        241 LTLTGLKLLYEKN 253


RESULT 5
AAU91170
ID   AAU91170 standard; Protein; 262 AA.
XX
AC   AAU91170;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #8.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus anthracis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
```

```
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54191.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 8; Page 98-99; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
     is the amino acid sequence of a pantothenate kinase (CoaX) protein
     described in the invention.
XX
SQ   Sequence   262 AA;

  Query Match             77.5%;  Score 1034;  DB 23;  Length 262;
  Best Local Similarity   75.1%;  Pred. No. 7.3e-108;
  Matches  190;  Conservative   34;  Mismatches   29;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :: |:|||||| |||| : :|:|  |||:|| |||||||:||::: | :| || || : ||
Db          1 MIFVLDVGNTNAVLGVFEEGELRQHWRMETDRHKTEDEYGMLVKQLLEHEGLSFEDVKGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |:|||||||||||||| ||| |:| :||||:||||||||:||:|||||||||||| |||
Db         61 IVSSVVPPIMFALERMCEKYFKIKPLVVGPGIKTGLNIKYENPREVGADRIVNAVAGIHL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              ||:|||:||||||||||||:| | |||| | ||| || |||||||||||||||:| :::
Db        121 YGSPLIIVDFGTATTYCYINEEKHYMGGVITPGIMISAEALYSRAAKLPRIEITKPSSVV 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              |||||||||||||:|||||||||||||| :|||: |||||||| ||: ||: ||:||||
Db        181 GKNTVSAMQSGILYGYVGQVEGIVKRMKEEAKQEPKVIATGGLAKLISEESNVIDVVDPF 240

Qy        241 LTLKGLELIYERN 253
              |||||| ::||||
Db        241 LTLKGLYMLYERN 253


RESULT 6
AAU91171
ID   AAU91171 standard; Protein; 254 AA.
XX
AC   AAU91171;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #9.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bacillus halodurans.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54192.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 10; Page 100; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   254 AA;

  Query Match             73.9%;  Score 987;  DB 23;  Length 254;
  Best Local Similarity   72.3%;  Pred. No. 1.4e-102;
  Matches  183;  Conservative   33;  Mismatches   37;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||||||||||||| |  | :|||: ||| |||||: | :||||||:|| |: ||||
Db          1 MILVIDVGNTNTVLGVYQDETLVHHWRLATSRQKTEDEYAMTVRSLFDHAGLQFQDIDGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              :|||||||:||:||:|| |||||: | |:|||:|||||||||||||||||||||||| |
Db         61 VISSVVPPMMFSLEQMCKKYFHVTPMIIGPGIKTGLNIKYDNPKEVGADRIVNAVAAIEL 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || | |||||||||||| |:| ||| || ||||| ||||||| ||:||||||| :|  ::
Db        121 YGYPAIVVDFGTATTYCLINEKKQYAGGVIAPGIMISTEALYHRASKLPRIEIAKPKQVV 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              | ||: :||||| :||| ||:|:||||| ||: : ||||||||| ||  ||: ||::| |
Db        181 GTNTIDSMQSGIFYGYVSQVDGVVKRMKAQAESEPKVIATGGLAKLIGTESETIDVIDSF 240

Qy        241 LTLKGLELIYERN 253
              ||||||:|||::|
Db        241 LTLKGLQLIYKKN 253


RESULT 7
ABB47661
ID   ABB47661 standard; Protein; 259 AA.
XX
AC   ABB47661;
XX
DT   05-FEB-2002  (first entry)
XX
DE   Listeria monocytogenes protein #365.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
```

```
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C,  Frangeul L,  Couve E,  Rusniok C,  Fsihi H,  Dehoux P;
PI   Dussurget O,  Chetouani F,  Nedjari H,  Glaser P,  Kunst F,  Cossart P;
PI   Daniels J,  Goebel W,  Kreft J,  Kuhn M,  Ng E,  Vazquez-Boland JÁ;
PI   Dominguez-Bernal G,  Garrido-Garcia P,  Tierrez-Martinez A,  Amend A;
PI   Chakraborty T,  Domann E,  Hain T,  Berche P,  Charbit A,  Durant L;
PI   Perez-Diaz J,  Baquero F,  Garcia Del Portillo F,  Gomez-Lopez N;
PI   Maduenio E,  De Pablos B,  Wehland J,  Kaerst U,  Entian K,  Hauf J;
PI   Rose M,  Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides  -
XX
PS   Claim 6; SEQ ID No 366; 192pp; French.
XX
CC   The present invention relates to the genome sequence of Listeria
CC   monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC   it are useful for selecting probes and primers for detecting genes in L.
CC   monocytogenes and related organisms, and for studying genetic
CC   polymorphisms and other genomes. The present sequence is a protein
CC   encoded by the genome sequence of the present invention. Proteins
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   259 AA;

  Query Match           63.8%;  Score 852;  DB 23;  Length 259;
  Best Local Similarity   63.4%;  Pred. No. 2.4e-87;
  Matches  161;  Conservative   37;  Mismatches   56;  Indels    0;  Gaps    0;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::|||||||||  :|||   ||  |||: | ||:| || || : : | :: |    | ||
Db          1 MILVIDVGNTNCTVGVYEKQKLLKHWRMTTDRHRTSDELGMTVLNFFSYANLTPSDIQGI 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||||| |:| || :||:| | |||||:|||||:| |||:|:|:|||||||||
Db         61 IISSVVPPIMHAMETMCVRYFNIRPLIVGPGIKTGLNLKVDNPREIGSDRIVNAVAASEE 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || |:|||||||||:|||||:  | |||||||| |||||||:|||||||::|     ||
Db        121 YGTPVIVVDFGTATTFCYIDESGVYQGGAIAPGIMISTEALYNRAAKLPRVDIAESSQII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
              ||:|||:||:|| :|:||| |||:  || |:     |:|||||| :|  :|  :|||:|||
Db        181 GKSTVSSMQAGIFYGFVGQCEGIIAEMKKQSNASPVVVATGGLARMITEKSSAVDILDPF 240

Qy        241 LTLKGLELIYERNR 254
              ||||||||:| ||:
Db        241 LTLKGLELLYRRNK 254


RESULT 8
AAU91175
ID   AAU91175 standard; Protein; 256 AA.
XX
AC   AAU91175;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #13.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Clostridium difficile.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54196.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 105; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   256 AA;

  Query Match           59.5%;  Score 794.5;  DB 23;  Length 256;
  Best Local Similarity   60.8%;  Pred. No. 7.2e-81;
  Matches  155;  Conservative   40;  Mismatches   59;  Indels    1;  Gaps    1;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :||| |||||| |||:|   ||  :|||:| | || ||:|::: :|||:  :    || :
Db          1 MLLVFDVGNTNMVLGIYKGDKLVNYWRIKTDREKTSDEYGILISNLFDYDNVNISDIDDV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| :| :||  | ||   :| |||||:||||||||||||:|||||||||| |
Db         61 IISSVVPNVMHSLENFCIKYCKKQPLIVGPGIKTGLNIKYDNPKQVGADRIVNAVAGIEK 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              || | |:||||||||:| | |  :|:|| ||||| ||:|||:  |:||||:|: :|   |
Db        121 YGAPSILVDFGTATTFCAISEKGEYLGGTIAPGIKISSEALFQSASKLPRVELAKPGMTI 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAK-QDLKVIATGGLAPLIANESDCIDIVDP 239
              |:|||||||||::|||| |: |:  || :    |:|||||||| |||:|:  || ||
Db        181 CKSTVSAMQSGIIYGYVGLVDKIISIMKKELNCDDVKVIATGGLAKLIASETKSIDYVDG 240

Qy        240 FLTLKGLELIYERNR 254
              ||||:|| :|||:|:
Db        241 FLTLEGLRIIYEKNQ 255


RESULT 9
AAU91154
ID   AAU91154 standard; Protein; 255 AA.
XX
```

AC AAU91154;
XX
DT 05-JUN-2002 (first entry)
XX
DE Geobacter sulfurreducens pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Geobacter sulfurreducens.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
(OMNI-) OMNIGENE BIOPRODUCTS INC.

PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 72-73; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 255 AA;

Query Match 56.6%; Score 756; DB 23; Length 255;
Best Local Similarity 57.5%; Pred. No. 1.6e-76;
Matches 146; Conservative 46; Mismatches 62; Indels 0; Gaps 0;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|||||||||| |||:|   :|   ||: | : :| ||:|:::  ||  :||  :||  :
Db          1 MLLVIDVGNTNIVLGIYDGERLVRDWRVSTDKARTTDEYGILINELFRLAGLGLDQIRAV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||||:   |||:   || : | :||||:|||: |:||||:||||||||||||
Db         61 IISSVVPPLTGVLERLSLGYFGMRPLVVGPGIKTGMPIQYDNPREVGADRIVNAVAGYEK 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              |   ||:||||||||: |::   :| |||||||: ||||||: ||:||||::| ||  ||
Db        121 YRTSLIIVDFGTATTFDYVNRKGEYCGGAIAPGLVISTEALFQRASKLPRVDIIRPSAII 180

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDPF 240
               :|||::||:|| :|||| |: || ||| ::|   :|||||||| ||| ||  |: |: :
Db        181 ARNTVNSMQAGIYYGYVGLVDEIVTRMKAESKDAPRVIATGGLASLIAPESKTIEAVEEY 240

Qy        241 LTLKGLELIYERNR 254
              |||:|| ::|||||
Db        241 LTLEGLRILYERNR 254
```

RESULT 10
AAU91151
ID AAU91151 standard; Protein; 265 AA.

XX
AC AAU91151;
XX
DT 05-JUN-2002 (first entry)
XX
DE Streptomyces coelicolor pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Streptomyces coelicolor.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 69-70; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 265 AA;

Query Match 49.1%; Score 655; DB 23; Length 265;
Best Local Similarity 51.0%; Pred. No. 4e-65;
Matches 133; Conservative 46; Mismatches 74; Indels 8; Gaps 3;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ---- 56
              :|| ||||||:||||::    :  |||| |   :| ||  ::|: |     |: ::
Db          1 MLLTIDVGNTHTVLGLFDGEDIVEHWRISTDSRRTADELAVLLQGLMGMHPLLGDELGDG 60

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              |||| | : || ::  |  :  :|:   | : | ||:|||: |  |:||||||||:|||
Db         61 IDGIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPILTDHPKEVGADRIINAV 120

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||: ||| | ||||||||||:  :    :|:|| ||||| || |||  : |:| :||: |
Db        121 AAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVAR 180

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQA---KQDLKVIATGGLAPLIANESD 232
              | ::|||||| ||||||::|: |||:|:| ||  :      |:|||||||||::  ||
Db        181 PRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGGLAPMVLGESS 240

Qy        233 CIDIVDPFLTLKGLELIYERN 253
               ||  :|:||| || |:|||
Db        241 VIDEHEPWLTLMGLRLVYERN 261
```

RESULT 11
AAU91150

ID AAU91150 standard; Protein; 250 AA.
XX
AC AAU91150;
XX
DT 05-JUN-2002 (first entry)
XX
DE Clostridium acetobutylicum pantothenate kinase CoaX.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Clostridium acetobutylicum.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 68-69; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 250 AA;

Query Match 48.4%; Score 646.5; DB 23; Length 250;
Best Local Similarity 53.9%; Pred. No. 3.4e-64;
Matches 125; Conservative 46; Mismatches 56; Indels 5; Gaps 1;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              ::||:|||||| |||:|:| ||   ||: | . :: ||:|: : :||    |    ::|:
Db         18 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRSADEYGIQVMNLFQQDKLDPTLVEGV 77

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              ||||||| ||::|| |  ||| | | :||||:|||:||||||||||||||||||||  :
Db         78 IISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGADRIVNAVAAHEI 137

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNII 180
              |   ||::|||||||:| : ||  |:|||| ||| :|:|||: :||||||:|: :|   |
Db        138 YKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKLPRVELIKPAYAI 197

Qy        181 GKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLK-----VIATGGLAPLI 227
               |||:|::||||:: |: ||: : :::|       :     |:||||| ||
Db        198 CKNTISSIQSGIVYRYLRQVKYLFEKLKENLPDGRRTRTSLVLATGGLAKLI 249
```

RESULT 12
AAU91173
ID AAU91173 standard; Protein; 260 AA.
XX
AC AAU91173;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #11.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Caulobacter crescentus.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54194.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 10; Page 102-103; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 260 AA;

Query Match 43.2%; Score 577; DB 23; Length 260;
Best Local Similarity 44.4%; Pred. No. 2.4e-56;
Matches 114; Conservative 50; Mismatches 91; Indels 2; Gaps 1;

```
Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|| |: |||||:  ::         ||  |   :| ||: : |  |    || |  || :
Db          1 MLLAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAV 60

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHL 120
              |||||||  :| |  :  :||::|| ::|   | |:::: : | | ||||:|||: |  :
Db         61 IISSVVPQSIFNLRNLSRRYFNVEPLVIGENAKLGIDVRIEKPSEAGADRLVNAIGAAMV 120

Qy        121 YGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRP--DN 178
              |  ||:|:| |||||:  :  :  : || ||||| :| :||:  |||||| | ||  :
Db        121 YPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQALHEAAAKLPRIAIQRPAGNR 180

Qy        179 IIGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVD 238
              |:| :|||||||: :||:  :||:| |:| :  : : ||||||:| |    :| ||  |
Db        181 IVGTDTVSAMQSGVFWGYISLIEGLVARIKAERGEPMTVIATGGVASLFEGATDSIDHFD 240

Qy        239 PFLTLKGLELIYERNRV 255
               ||::||  || || :
Db        241 SDLTIRGLLEIYRRNTI 257
```

RESULT 13
AAU91153
ID AAU91153 standard; Protein; 258 AA.

```
XX
AC   AAU91153;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Rhodobacter capsulatus pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Rhodobacter capsulatus.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
     (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 71-72; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   258 AA;

  Query Match             40.2%;  Score 537;  DB 23;  Length 258;
  Best Local Similarity   43.9%;  Pred. No. 7.8e-52;
  Matches  112;  Conservative   46;  Mismatches   95;  Indels    2;  Gaps    2;

Qy          1 LLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60
              :|| || |||||| |:        ||| |   :| ||: : | :|    ||   :|
Db          1 MLLCIDCGNTNTVFSVWDGTDFAATWRIATDHRRTADEYFVWLNTLMQLKGLQ-GRISEA 59

Qy         61 IISSVVPPIMFALERMCTKYFHIEPQIVG-PGMKTGLNIKYDNPKEVGADRIVNAVAAIH 119
              ||||  | ::| |  :| :||   | :|| || :  :  : |    || ||:|| ||
Db         60 IISSTAPRVVFNLRVLCNRYFDCRPYVVGKPGCELPVAPRVDPGTTVGPDRLVNTVAGYD 119

Qy        120 LYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNI 179
               :|  ||||||||||||:  :  :  |:|| |||| :| |||:  || || :::|:|  :
Db        120 RHGGDLIVVDFGTATTFDVVAPDGAYIGGVIAPGVNLSLEALHMAAAALPHVDVTKPQGV 179

Qy        180 IGKNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQDLKVIATGGLAPLIANESDCIDIVDP 239
              || |||: :|||: :||:| |||||:::: :  : :|||||||| |     |  | |:
Db        180 IGTNTVACIQSGVYWGYIGLVEGIVRQIRMERDRPMKVIATGGLASLFDLGFDLFDKVED 239

Qy        240 FLTLKGLELIYERNR 254
               ||: || ||:: |:
Db        240 DLTMHGLRLIFDYNK 254


RESULT 14
AAU91176
ID   AAU91176 standard; Protein; 219 AA.
XX
AC   AAU91176;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #14.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Dehalococcoides ethenogenes.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54197.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 106-107; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   219 AA;

  Query Match             38.6%;  Score 515;  DB 23;  Length 219;
  Best Local Similarity   45.8%;  Pred. No. 1.9e-49;
  Matches   97;  Conservative   45;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          2 LLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGII 61
              |: :|:|||:  :|::   ||  :| : :   :  ||:  :|  |  |:|:  |::: :|
Db          5 LVAVDIGNTSVNIGIFEGEKLLANWHLGSVAQRMADEYASLLLGLLQHAGIHPEELNRVI 64

Qy         62 ISSVVPPIMFALERMCTKYFHIEPQIVGPGMKTGLNIKYDNPKEVGADRIVNAVAAIHLY 121
              : |||||:    | :   ||   | :|| |:|:|: :: |||:||||||||| ||  ||
Db         65 MCSVVPPLTTTFEEVFKSYFKAAPLVVGAGIKSGVKVRMDNPREVGADRIVNAAAARVLY 124

Qy        122 GNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITRPDNIIG 181
                  |:|| |||||:  : |   |:||||||||  | :|:  : :|||:||| ||  :||
Db        125 PGACIIVDMGTATTFDTLSEGGAYIGGAIAPGIATSAQAIAEKTSKLPKIEIIRPAKVIG 184

Qy        182 KNTVSAMQSGILFGYVGQVEGIVKRMKWQAKQ 213
               |||||||||| |||:| || :|:|:: :  |
Db        185 SNTVSAMQSGIYFGYIGLVEELVRRIQTELGQ 216


RESULT 15
AAU91152
ID   AAU91152 standard; Protein; 272 AA.
XX
```

```
AC   AAU91152;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Mycobacterium tuberculosis pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 10; Page 70-71; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   272 AA;

  Query Match             37.0%;  Score 494.5;  DB 23;  Length 272;
  Best Local Similarity   38.8%;  Pred. No. 5.3e-47;
  Matches  102;  Conservative   55;  Mismatches   93;  Indels   13;  Gaps    4;

Qy          1 LLLVIDVGNTNTVLGVY----HDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQ 56
              :|| ||| ||:||:|:        |:   ||| |    | ||  : :  |        |:
Db          1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQQWRIRTESEVTADELALTIDGLIGEDS---ER 57

Qy         57 IDGIIISSVVPPIMFALERMCTKYFHIEPQI-VGPGMKTGLNIKYDNPKEVGADRIVNAV 115
              : |    | || ::  :  |  :|:   | : : ||::||: :  ||||||||||||| :
Db         58 LTGTAALSTVPSVLHEVRIMLDQYWPSVPHVLIEPGVRTGIPLLVDNPKEVGADRIVNCL 117

Qy        116 AAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEALYSRAAKLPRIEITR 175
              ||   :    ||||||::     :    :::|||||||: :|:::|  :|:| | |:|: |
Db        118 AAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAARSAALRRVELAR 177

Qy        176 PDNIIGKNTVSAMQSGILFGYVGQVEGIVKRMK-----WQAKQDLKVIATGGLAPLIANE 230
              | :::|||||  ||:| :||: | |:|:| |::     :    |: ::|||  |||:  |
Db        178 PRSVVGKNTVECMQAGAVFGFAGLVDGLVGRIREDVSGFSVDHDVAIVATGHTAPLLLPE 237

Qy        231 SDCIDIVDPFLTLKGLELIYERN 253
                :|  |  |||:|| |::|||
Db        238 LHTVDHYDQHLTLQGLRLVFERN 260


Search completed: June 24, 2003, 21:46:10
Job time : 27.4883 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 25.4617 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-20
Perfect score: 1248
Sequence: 1 MILELDCGNSLIKWRVIEGA.....GARIMPDLVFVGLALACPIE 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1248	100.0	248	23	AAU91164	Pantothenate kinas
2	794.5	63.7	249	23	AAU91178	Pantothenate kinas
3	789.5	63.3	249	23	AAU91182	Pantothenate kinas
4	236	18.9	389	21	AAY74909	Neisseria meningit
5	234.5	18.8	592	20	AAY38615	Neisseria meningit
6	234.5	18.8	592	21	AAY74912	Neisseria meningit
7	234.5	18.8	592	23	AAU91169	Pantothenate kinas
8	232.5	18.6	455	20	AAY38617	Neisseria gonorrho
9	232.5	18.6	455	21	AAY74908	Neisseria gonorrhe
10	232.5	18.6	460	23	AAU91167	Pantothenate kinas
11	232.5	18.6	592	20	AAY38618	Neisseria gonorrho
12	232.5	18.6	592	21	AAY74911	Neisseria gonorrhe
13	231.5	18.5	455	21	AAY74910	Neisseria meningit
14	231.5	18.5	592	20	AAY38616	Neisseria meningit
15	231.5	18.5	592	21	AAY74913	Neisseria meningit
16	231.5	18.5	592	23	AAU91166	Pantothenate kinas
17	195.5	15.7	244	23	AAU91168	Pantothenate kinas
18	193	15.5	242	23	AAU91180	Pantothenate kinas
19	185.5	14.9	189	20	AAY38614	Neisseria meningit
20	178.5	14.3	267	23	AAU91162	Bordella pertussis
21	160	12.8	260	23	AAU91173	Pantothenate kinas
22	159.5	12.8	223	19	AAW98422	H. pylori GHPO 344
23	159.5	12.8	223	23	AAU91161	Helicobacter pylor
24	159.5	12.8	223	23	AAU91181	Pantothenate kinas
25	159.5	12.8	250	23	AAU91150	Clostridium acetob
26	157	12.6	257	23	AAU91174	Pantothenate kinas
27	149	11.9	272	23	AAU91152	Mycobacterium tube
28	147	11.8	272	22	AAG81225	Mycobacterium tube
29	145	11.6	246	23	AAU91156	Thermotoga maritim
30	144.5	11.6	241	23	AAU91179	Pantothenate kinas
31	141	11.3	255	23	AAU91154	Geobacter sulfurre
32	138.5	11.1	265	23	AAU91151	Streptomyces coeli
33	136	10.9	219	23	AAU91176	Pantothenate kinas
34	136	10.9	258	23	AAU91153	Rhodobacter capsul
35	136	10.9	262	23	AAU91170	Pantothenate kinas
36	136	10.9	273	23	AAU91157	Treponema pallidum
37	133	10.7	258	22	AAU01243	B. subtilis novel
38	133	10.7	258	23	AAU91149	Bacillus subtilis
39	131.5	10.5	259	23	ABB47661	Listeria monocytog
40	123.5	9.9	258	23	AAU91172	Pantothenate kinas
41	121	9.7	254	23	AAU91171	Pantothenate kinas
42	120.5	9.7	209	23	AAU91165	Pantothenate kinas
43	119.5	9.6	256	23	AAU91175	Pantothenate kinas
44	111	8.9	233	23	AAU91163	Pantothenate kinas
45	107	8.6	212	23	AAU91177	Pantothenate kinas

ALIGNMENTS

RESULT 1
AAU91164
ID AAU91164 standard; Protein; 248 AA.
XX
AC AAU91164;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #2.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -

```
XX
PS   Claim 6; Page 83-84; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   248 AA;

  Query Match            100.0%;  Score 1248;  DB 23;  Length 248;
  Best Local Similarity  100.0%;  Pred. No. 4.2e-136;
  Matches  248;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60

Qy         61 TSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVI 120

Qy        121 DLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQAT 180

Qy        181 AEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVG 240

Qy        241 LALACPIE 248
              ||||||||
Db        241 LALACPIE 248


RESULT 2
AAU91178
ID   AAU91178 standard; Protein; 249 AA.
XX
AC   AAU91178;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #16.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas putida.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54199.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 10; Page 108-109; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match            63.7%;  Score 794.5;  DB 23;  Length 249;
  Best Local Similarity  63.5%;  Pred. No. 1.7e-83;
  Matches  158;  Conservative   31;  Mismatches   59;  Indels    1;  Gaps    1;

Qy          1 MILELDCGNSLIKWRVIEGAARSV-AGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQ 59
              |||||||||| ||||||  |   :  ||: :|| ||| :: :  :: :  ||:|||||||:
Db          1 MILELDCGNSFIKWRVIHVADAVIEGGGIVDSDQALVAEVAALASVRLTGCRIVSVRSEE 60

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              ||  | | : | | | | ||   :::|||||||| ||||||:||||| : | |||| ||||
Db         61 ETDALCALIAQAFAVQAKVAHPVREMAGVRNGYDDYQRLGMDRWLAALGAFHLAKGACLV 120

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQA 179
              ||||||  :| |:||| |||||||||| ||||||||||||||||| | |||:|| ||::
Db        121 IDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRIRYDDASAERALSSLSPGRS 180

Qy        180 TAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFV 239
              | |||||||:|||:||   |   |  | | :  :|||||||  |||  |  ||::||||||
Db        181 TVEAVERGCVLMLQGFAYTQLEQARVLWGEEFTVFLTGGDAPLVRAALPQARVVPDLVFV 240

Qy        240 GLALACPIE 248
              |||:|||::
Db        241 GLAMACPLD 249


RESULT 3
AAU91182
ID   AAU91182 standard; Protein; 249 AA.
XX
AC   AAU91182;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #20.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Pseudomonas syringae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54203.
```

```
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 10; Page 113-114; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   249 AA;

  Query Match             63.3%;  Score 789.5;  DB 23;  Length 249;
  Best Local Similarity   64.5%;  Pred. No. 6.4e-83;
  Matches  160;  Conservative   26;  Mismatches   61;  Indels    1;  Gaps    1;

Qy          1 MILELDCGNSLIKWRVI-EGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQ 59
              |||||||||| ||||:| :  :  |:||: :|| ||:| | :        ||||||| :
Db          1 MILELDCGNSFIKWRIITKSCSTLVSGGVVDSDTALLECLGNLSGAAFSDCRLVSVRSAE 60

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              ||::||  |   | :|: : |:   :|||| ||| |: |||||||| | |:|| | ||||
Db         61 ETAKLVCALADTFSISPVCAAPAPELAGVINGYDDFARLGLDRWLAFVGAYHLVKGACLV 120

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQA 179
              ||||||:||| | | | ||||:||||| |||:|||||||||||||||| :||  | ||:|
Db        121 IDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEAEKALVRLVPGRA 180

Qy        180 TAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFV 239
              |||||||||| ||||||   | :|: |  | |  ||:||||| || | | ||||:||||||
Db        181 TAEAVERGCSLMLRGFAMTQIEIAREYWGDDFAIFVTGGDAVLVADVLPGARIVPDLVFV 240

Qy        240 GLALACPI 247
              ||||||||
Db        241 GLALACPL 248


RESULT 4
AAY74909
ID   AAY74909 standard; Protein; 389 AA.
XX
AC   AAY74909;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53671.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 710; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   389 AA;

  Query Match             18.9%;  Score 236;  DB 21;  Length 389;
  Best Local Similarity   28.3%;  Pred. No. 2e-18;
  Matches   73;  Conservative   39;  Mismatches  108;  Indels   38;  Gaps    8;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLV-SVRSE 58
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :    :
Db        138 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 197

Qy         59 QETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACL 118
              |   ||  ::| |          || : | |:|| |    :  | |||   : :   :: ||:
Db        198 QVQEQLARKIEWL-------PSSAQALFGIRNHYRHPEEHGSDRWFNALGSRRFSRNACV 250

Qy        119 VIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP-- 176
              |:  ||||| | :  || :||| | ||  ||:  |    |       |   |       |
Db        251 VVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFP 303

Qy        177 ---GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA 230
                 | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        304 TTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPA 358

Qy        231 -------RIMPDLVFVGL 241
                    |:  :||  ||
Db        359 FLAENTVRVADNLVIYGL 376


RESULT 5
AAY38615
ID   AAY38615 standard; Protein; 592 AA.
XX
AC   AAY38615;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
```

```
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             18.8%;  Score 234.5;  DB 20;  Length 592;
  Best Local Similarity   27.6%;  Pred. No. 5.5e-18;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | |  :     | ::|: | ::|:         |   :: :||| |  | : |  |
Db        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561

Qy        231 ------RIMPDLVFVGL 241
                   |:  :||  ||
Db        562 LAENTVRVADNLVIYGL 578


RESULT 6
AAY74912
ID   AAY74912 standard; Protein; 592 AA.
XX
AC   AAY74912;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53674.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 714; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             18.8%;  Score 234.5;  DB 21;  Length 592;
  Best Local Similarity   27.6%;  Pred. No. 5.5e-18;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | |  :     | ::|: | ::|:         |   :: :||| |  | : |  |
Db        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561

Qy        231 ------RIMPDLVFVGL 241
```

```
                    |:  :||  ||
Db        562 LAENTVRVADNLVIYGL 578
```

RESULT 7
AAU91169
ID AAU91169 standard; Protein; 592 AA.
XX
AC AAU91169;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #7.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria meningitidis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54190.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 96-98; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 592 AA;

```
  Query Match             18.8%;  Score 234.5;  DB 23;  Length 592;
  Best Local Similarity   27.6%;  Pred. No. 5.5e-18;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |  |       |  |      |
Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561
```

Qy 231 ------RIMPDLVFVGL 241
 |: :|| ||
Db 562 LAENTVRVADNLVIYGL 578

RESULT 8
AAY38617
ID AAY38617 standard; Protein; 455 AA.
XX
AC AAY38617;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigen encoded by partial ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI; 1999-327407/27.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 175; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 455 AA;

```
  Query Match             18.6%;  Score 232.5;  DB 20;  Length 455;
  Best Local Similarity   27.2%;  Pred. No. 6.4e-18;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | |: ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        204 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 263

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        264 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 316

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |  |       |  |      |
Db        317 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 369
```

```
Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
              | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        370 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 424

Qy        231 ------RIMPDLVFVGL 241
                   |:  :||  ||
Db        425 LAENTVRVADNLVIHGL 441
```

RESULT 9
AAY74908
ID AAY74908 standard; Protein; 455 AA.
XX
AC AAY74908;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1290.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53670.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 709; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 455 AA;

```
  Query Match             18.6%;  Score 232.5;  DB 21;  Length 455;
  Best Local Similarity   27.2%;  Pred. No. 6.4e-18;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | |: ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        204 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 263

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||        |  |:|| |   :  | |||   : :   :: ||:|
Db        264 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 316

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        317 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 369

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        370 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 424

Qy        231 ------RIMPDLVFVGL 241
                   |:  :||  ||
Db        425 LAENTVRVADNLVIHGL 441
```

RESULT 10
AAU91167
ID AAU91167 standard; Protein; 460 AA.
XX
AC AAU91167;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #5.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54188.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 8; Page 92-94; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 460 AA;

```
  Query Match             18.6%;  Score 232.5;  DB 23;  Length 460;
  Best Local Similarity   27.2%;  Pred. No. 6.5e-18;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | |: ||| :|| :|    :|     | : :    :::|   : :  | :     :
Db        209 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 268

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        269 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 321

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        322 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 374

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        375 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 429

Qy        231 ------RIMPDLVFVGL 241
                    |:  :||  ||
Db        430 LAENTVRVADNLVIHGL 446


RESULT 11
AAY38618
ID   AAY38618 standard; Protein; 592 AA.
XX
AC   AAY38618;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigenic protein encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12075.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             18.6%;  Score 232.5;  DB 20;  Length 592;
  Best Local Similarity   27.2%;  Pred. No. 9.3e-18;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | |: ||| :|| :|    :|     | : :    :::|   : :  | :     :
Db        341 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        401 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        507 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 561

Qy        231 ------RIMPDLVFVGL 241
                    |:  :||  ||
Db        562 LAENTVRVADNLVIHGL 578


RESULT 12
AAY74911
ID   AAY74911 standard; Protein; 592 AA.
XX
AC   AAY74911;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1296.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53673.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
```

```
PS   Claim 2; Page 713; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             18.6%;  Score 232.5;  DB 21;  Length 592;
  Best Local Similarity   27.2%;  Pred. No. 9.3e-18;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | |: ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        341 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db        401 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        507 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 561

Qy        231 ------RIMPDLVFVGL 241
                   |:  :||  ||
Db        562 LAENTVRVADNLVIHGL 578


RESULT 13
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             18.5%;  Score 231.5;  DB 21;  Length 455;
  Best Local Similarity   27.7%;  Pred. No. 8.3e-18;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56
              | || ||| :||  :|    :  |     |   :  | :: |      : :  | :
Db        204 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 260

Qy         57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116
               : :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: |
Db        261 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 313

Qy        117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176
              |:|:  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        314 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 366

Qy        177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228
                    | | |  :     | ::|: | ::|:        |   :: :||| |  | : |
Db        367 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 421

Qy        229 GA-------RIMPDLVFVGL 241
                       |:  :||  ||
Db        422 PAFLAENTVRVADNLVIHGL 441


RESULT 14
AAY38616
ID   AAY38616 standard; Protein; 592 AA.
XX
AC   AAY38616;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
```

```
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             18.5%;  Score 231.5;  DB 20;  Length 592;
  Best Local Similarity   27.7%;  Pred. No. 1.2e-17;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56
              | || ||| :||  :|    :  |     |   :  | :: |      : :  | :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 397

Qy         57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116
               :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: |
Db        398 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 450

Qy        117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176
              |:|:  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        451 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 503

Qy        177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228
                  | | |  :     | ::|: | ::|:        |   :: :||| |  | : |
Db        504 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 558

Qy        229 GA-------RIMPDLVFVGL 241
               |       |:  :||  ||
Db        559 PAFLAENTVRVADNLVIHGL 578


RESULT 15
AAY74913
ID   AAY74913 standard; Protein; 592 AA.
XX
AC   AAY74913;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1300.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53675.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 715; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             18.5%;  Score 231.5;  DB 21;  Length 592;
  Best Local Similarity   27.7%;  Pred. No. 1.2e-17;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56
              | || ||| :||  :|    :  |     |   :  | :: |      : :  | :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 397

Qy         57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116
               :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: |
Db        398 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 450

Qy        117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176
              |:|:  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        451 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 503

Qy        177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228
                  | | |  :     | ::|: | ::|:        |   :: :||| |  | : |
Db        504 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 558

Qy        229 GA-------RIMPDLVFVGL 241
               |       |:  :||  ||
Db        559 PAFLAENTVRVADNLVIHGL 578
```

Search completed: June 24, 2003, 21:46:23
Job time : 25.4617 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 9.04559 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-20
Perfect score: 1248
Sequence: 1 MILELDCGNSLIKWRVIEGA.....GARIMPDLVFVGLALACPIE 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1248	100.0	248	2	H83111	hypothetical prote
2	234.5	18.8	592	2	B81009	BirA protein/Bvg a
3	231.5	18.5	592	2	H82031	probable biotin-[a
4	193	15.5	242	2	A82637	conserved hypothet
5	178.5	14.3	267	2	I40327	baf protein - Bord
6	176.5	14.1	273	2	E97293	probable transcrip
7	172.5	13.8	276	2	AI2292	hypothetical prote
8	160	12.8	261	2	B87489	transcription acti
9	159.5	12.8	223	2	F64627	hypothetical prote
10	158.5	12.7	223	2	G71887	hypothetical prote
11	149	11.9	272	2	A70955	hypothetical prote
12	145	11.6	246	2	D72320	conserved hypothet
13	138.5	11.1	265	2	T36391	hypothetical prote
14	137.5	11.0	274	2	H86937	conserved hypothet
15	136	10.9	273	2	D71326	conserved hypothet
16	131.5	10.5	259	2	AF1102	conserved hypothet
17	126.5	10.1	259	2	AF1464	conserved hypothet
18	121	9.7	254	2	F83660	hypothetical prote
19	120.5	9.7	209	2	H81382	hypothetical prote
20	111	8.9	233	2	S66100	conserved hypothet
21	107	8.6	257	2	S75559	hypothetical prote
22	101.5	8.1	451	2	T49021	hypothetical prote
23	98.5	7.9	229	2	E70465	hypothetical prote
24	97	7.8	728	2	C75601	cation-transportin
25	96.5	7.7	509	2	AC1667	Bifunctional phosp
26	92	7.4	224	2	A99571	conserved hypothet
27	91.5	7.3	822	2	E75523	ATP-dependent heli
28	90.5	7.3	940	2	A87502	hypothetical prote
29	89.5	7.2	428	2	T50731	protochlorophyllid
30	89.5	7.2	509	2	AE1295	Bifunctional phosp
31	89	7.1	324	2	S05396	bifunctional cycla
32	89	7.1	419	2	S75647	glutamate N-acetyl
33	89	7.1	908	2	AI0327	two-component regu
34	88.5	7.1	510	2	AI2012	hypothetical prote
35	88	7.1	487	2	D95887	probable aldehyde
36	87.5	7.0	361	2	H87700	citrate synthase [
37	87.5	7.0	517	2	G75593	uroporphyrin-III C
38	87.5	7.0	4151	2	G70944	probable polyketid
39	86.5	6.9	424	2	G98344	hypothetical prote
40	86.5	6.9	424	2	AH2937	hypothetical prote
41	86.5	6.9	432	2	E82300	glutamate-1-semial
42	86.5	6.9	564	2	S29246	urocanate hydratas
43	85.5	6.9	293	2	C98115	conserved hypothet
44	85.5	6.9	310	2	G75615	hypothetical prote
45	85.5	6.9	524	2	F97397	aarF protein (AF04

ALIGNMENTS

RESULT 1
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

Query Match 100.0%; Score 1248; DB 2; Length 248;
Best Local Similarity 100.0%; Pred. No. 3.8e-101;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60

Qy         61 TSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVI 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVI 120

Qy        121 DLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 DLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQAT 180

Qy        181 AEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVG 240

Qy        241 LALACPIE 248
              ||||||||
Db        241 LALACPIE 248
```

RESULT 2
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (stra
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009

```
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

  Query Match           18.8%;  Score 234.5;  DB 2;  Length 592;
  Best Local Similarity  27.6%;  Pred. No. 1.5e-12;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59
              | ||. ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119
              :  : :||  :  | ||       | :|| |   :  | |||   : :   :: ||:|
Db        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230
                | | |  :     | ::|: | ::|:       |   :: :||| |  | : |  |
Db        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561

Qy        231 ------RIMPDLVFVGL 241
                   |:  :||  ||
Db        562 LAENTVRVADNLVIYGL 578


RESULT 3
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidis (
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C;Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737911
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

  Query Match           18.5%;  Score 231.5;  DB 2;  Length 592;
  Best Local Similarity  27.7%;  Pred. No. 2.8e-12;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56
              | || ||| :||  :|   :  |     |   :  | :: |      : :  | :
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 397

Qy         57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116
                : :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: |
Db        398 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 450

Qy        117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176
              |:|:  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        451 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 503

Qy        177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228
                    | | |  :     | ::|: | ::|:       |   :: :||| |  | : |
Db        504 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 558

Qy        229 GA-------RIMPDLVFVGL 241
                |       |:  :||  ||
Db        559 PAFLAENTVRVADNLVIHGL 578


RESULT 4
A82637
conserved hypothetical protein XF1795 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84603.1; GSPDB:G
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; F
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; L
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martin
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1795

  Query Match           15.5%;  Score 193;  DB 2;  Length 242;
  Best Local Similarity  31.0%;  Pred. No. 2.2e-09;
  Matches   76;  Conservative   34;  Mismatches  111;  Indels   24;  Gaps    8;

Qy          6 DCGNSLIKWRVI-EGAAR--SVAGGLAESDDALVEQLTSQQALP-VRACRLVSVRSEQET 61
              | |||  |   : ||     :|   | |: ||      : | ||  |   | || :   |
Db          7 DLGNSRFKCASLREGVIGPVTVLPYLTETMDAF-----ALQELPRGRVAYLASVAAPAIT 61

Qy         62 SQLVARLEQLF---PVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACL 118
              : ::  |:  |    |:| ||:     ||||  |   :| |:||:|||: ::   :   |
Db         62 THVLEVLKIHFEQVQVAATVAA----CAGVRIAYAHPERFGVDRFLALLGSY--GEGNVL 115

Qy        119 VIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQ 178
              |: :|||:| ||:||:| |||| |    ||||  |      ::              |:  :
Db        116 VVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN------YLEFAE 169

Qy        179 ATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVF 238
                | :|:  ||       :      | : |     : | ||    :   |       | ||
Db        170 DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVL 229

Qy        239 VGLAL 243
               |||:
Db        230 DGLAI 234
```

```
RESULT 5
I40327
baf protein - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40327; S70669
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
J. Bacteriol. 177, 3801-3807, 1995
A;Title: Identification of a Bordetella pertussis regulatory factor required for transcr
A;Reference number: I40327; MUID:95325323; PMID:7601846
A;Accession: I40327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229
R;Allen, A.; Maskell, D.
Mol. Microbiol. 19, 37-52, 1996
A;Title: The identification, cloning and mutagenesis of a genetic locus required for lip
A;Reference number: S70669; MUID:96419162; PMID:8821935
A;Accession: S70669
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 239-267 <ALL>
A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: baf

  Query Match             14.3%;  Score 178.5;  DB 2;  Length 267;
  Best Local Similarity   29.6%;  Pred. No. 4.6e-08;
  Matches   75;  Conservative   32;  Mismatches   93;  Indels   53;  Gaps   11;

Qy          1 MILELDCGNSLIK--WRVIEG--AARSVAGGLAES--DDALVEQLTSQQALPVRACRLVS 54
              ||: :| ||| :|  |   :   |||  |    ::   |||   | :   || |  | :
Db          1 MIILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLAT---LPRRPQRALG 57

Qy         55 VRSEQETSQLVARLEQLFPVSALVASSGKQL---------AGVRNGYLDYQRLGLDRWLA 105
              |         || | :   ::| : : |  :          |:|||| :  :|| |||
Db         58 VN--------VAGLARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWAC 109

Qy        106 LVAAHHLAKK-----ACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRI 160
              :|    ||::       ||   ||| | | :  | |  || | ||  :|| |   |  :
Db        110 MVGV--LARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHL 167

Qy        161 RYDDA-------EARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLGPDCEI 213
                |       :  :|:||   | | |:|         | :  |:    :  |   ||
Db        168 PLADGLVADYPIDTHQAIAS---GIAAAQA----------GAIVRQWLAGRQRYGQAPEI 214

Qy        214 FLTGGDAELVRDE 226
              :: ||    || |
Db        215 YVAGGGWPEVRQE 227


RESULT 6
E97293
probable transcription regulator, homolog of Bvg accessory factor [imported] - Clostridi
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C;Accession: E97293
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: E97293
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-273 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK81136.1; PID:g15026270; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC3200
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             14.1%;  Score 176.5;  DB 2;  Length 273;
  Best Local Similarity   26.5%;  Pred. No. 7e-08;
  Matches   71;  Conservative   43;  Mismatches  113;  Indels   41;  Gaps    9;

Qy          1 MILELDCGNSLI-----------KWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVR 48
              :|| || ||: |            :||:     ||      |:     |  |  |  |
Db         12 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRS-----ADEYGIQVMNLFQQDKLDPT 66

Qy         49 ACRLVSVRSEQETSQLVARLEQL----FPVSALVASSGKQLAGVRNGYLDYQRLGLDRWL 104
                | : |      ::  || :    | :: ||   | :  |:   | : : :| || :
Db         67 LVEGVIISS--VVPNIMYSLEHMIRKYFKINPLVVGPGIK-TGINIKYDNPKEVGADRIV 123

Qy        105 ALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDD 164
              |||| : |:: ::|| ||| |   |  :| :||| |||: :    |        ::
Db        124 NAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEAL--------FEK 175

Qy        165 AEARRALASLQPGQA----TAEAVERGCLLMLRGFVREQYAMACELLGPDCE----IFLT 216
              |     :  ::| |    | ::: | :   | ||      | |  : |    :  |
Db        176 AAKLPRVELIKPAYAICKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVAT 235

Qy        217 GGDAELVRDELAGARIM-PDLVFVGLAL 243
              || |:|: :|     :: | |   || :
Db        236 GGLAKLISEEAKNVDVINPFLTLEGLRI 263


RESULT 7
AI2292
hypothetical protein alr3896 [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp.
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C;Accession: AI2292
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iri
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Taba
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacteriu
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AI2292
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-276 <KUR>
A;Cross-references: GB:BA000019; PIDN:BAB75595.1; PID:g17133030; GSPDB:GN00179
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: alr3896

  Query Match             13.8%;  Score 172.5;  DB 2;  Length 276;
  Best Local Similarity   27.1%;  Pred. No. 1.6e-07;
  Matches   76;  Conservative   43;  Mismatches   94;  Indels   67;  Gaps   12;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQETS 62
              | |: ||| : | :  | :           :::::|| : :         : | ||:
Db         15 LALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGET-------KLEVGSEE--- 64

Qy         63 QLVARLEQLF-----------PVSALVASSGKQLAGVRNGYLD--------------YQR 97
                 :::|            |:   :||   |   :   ||:               |
Db         65 ------KEIFFTFFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPT 118

Qy         98 LGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVH--LGGYICPGMTLMRSQLRT 155
              ||:|| |||  |         |||| |||:|    ||||    :|| | ||: |  : |
Db        119 LGIDRALALWGAGMSWGFPVLVIDAGTALT--FTAADGGKNLVGGAILPGVGLQFASLGQ 176

Qy        156 HTRRIRYDDAEARRALASLQPGQA--TAEAVERGCLLML----RGFVREQYAMACELLGP 209
               | ::   : |   |: || |  |  | ||:: | :  |     | |  |  :     | |
Db        177 QTGQLPQVEME---AIKSLPPRFALNTTEAIQSGVIYTLIAGMRDFTEEWLS-----LFP 228

Qy        210 DCEIFLTGGDAELVRDEL------AGARIM--PDLVFVGL 241
              | :: : |||  |: : |        ||::  |:|:| |:
Db        229 DGKVAIKGGDRILLLNYLQALYPDLAARLIVEPNLIFWGM 268
```

```
RESULT 8
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             12.8%;  Score 160;  DB 2;  Length 261;
  Best Local Similarity   28.3%;  Pred. No. 1.8e-06;
  Matches   67;  Conservative   30;  Mismatches  116;  Indels   24;  Gaps    8;

Qy          1 MILELDCGNSLIKWRVIEGAA-----RSVAGGLAESDDALV--EQLTSQQALPVRACRLV 53
              |:| :: ||:   : : :||:     ||       :|: :|   || | | |  ||   |
Db          2 MLLAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAV 61

Qy         54 SVRSEQETSQLVARLEQL----FPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAA 109
               : |     | :  | |    | |  ||     :| |:           | || :  : |
Db         62 IISS--VVPQSIFNLRNLSRRYFNVEPLVIGENAKL-GIDVRIEKPSEAGADRLVNAIGA 118

Qy        110 HHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARR 169
                :     :||| ||| | |:|||||   || | ||: |    |  |    :      :|
Db        119 AMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQAL--HEAAAKLPRIAIQR 176

Qy        170 ALASLQPGQATAEAVERGC----LLMLRGFVREQYAMACELLGPDCEIFLTGGDAEL 222
                  :  | |  |:: |     : :: | |   |      |    :  ||| | |
Db        177 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAER----GEPMTVIATGGVASL 229


RESULT 9
F64627
hypothetical protein HP0862 - Helicobacter pylori (strain 26695)
C;Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 18-Aug-2000
C;Accession: F64627
R;Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.
 Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A;Reference number: A64520; MUID:97394467; PMID:9252185
A;Accession: F64627
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-223 <TOM>
A;Cross-references: GB:AE000596; GB:AE000511; NID:g2313982; PIDN:AAD07916.1; PID:g231400
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

  Query Match             12.8%;  Score 159.5;  DB 2;  Length 223;
  Best Local Similarity   28.6%;  Pred. No. 1.7e-06;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112
              :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::
Db         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97

Qy        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172
                 :|:| |:|:| ||:  :| |||| | ||:            :|  |  : |
Db         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139

Qy        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221
                 || :|          | :||  | :| :          :|| : |  : :|:| ||||:
Db        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190

Qy        222 LVRDELAGARIMPDLVFVGLALA 244
                :   |  :     ||| |: :|
Db        191 YLSAFLPHSVCKERLVFDGMEIA 213


RESULT 10
G71887
hypothetical protein jhp0796 - Helicobacter pylori (strain J99)
C;Species: Helicobacter pylori
A;Variety: strain J99
C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 18-Aug-2000
C;Accession: G71887
R;Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D
; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.
Nature 397, 176-180, 1999
A;Title: Genomic sequence comparison of two unrelated isolates of the human gastric
A;Reference number: A71800; MUID:99120557; PMID:9923682
A;Accession: G71887
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-223 <ARN>
A;Cross-references: GB:AE001509; GB:AE001439; NID:g4155350; PIDN:AAD06372.1; PID:g41
A;Experimental source: strain J99
C;Genetics:
A;Gene: jhp0796
C;Superfamily: Campylobacter jejuni hypothetical protein Cj0394c

  Query Match             12.7%;  Score 158.5;  DB 2;  Length 223;
  Best Local Similarity   28.6%;  Pred. No. 2e-06;
  Matches   58;  Conservative   33;  Mismatches   61;  Indels   51;  Gaps    8;

Qy         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112
              :||  | | : |       :|       : | :||  :   ||  ||:|| :| :|  :
Db         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYIGLGIDRQMACLAVVN- 97

Qy        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172
                 :::| |:|:| |||  :| |||| | ||:            :|  |  : |
Db         98 ----GVIVDAGSAITIDLV-KEGKHLGGCILPGLA------------QYVHAYKKSAKI 139

Qy        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221
                 || :|          | :||  | :| :          ::| : |  | :|:| ||||:
Db        140 LEQPFKALDSLEVLPKNTRDAVNYGMILSI---------ISCIQHLAKDQKIYLCGGDAK 190

Qy        222 LVRDELAGARIMPDLVFVGLALA 244
                :   |  :     ||| |: :|
Db        191 YLSAFLPHSVCKERLVFDGMEIA 213


RESULT 11
A70955
hypothetical protein Rv3600c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70955
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordo
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70955
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-272 <COL>
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08944.1; PID:g2113
A;Experimental source: strain H37Rv
C;Genetics:
```

```
A;Gene: Rv3600c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match           11.9%;  Score 149;  DB 2;  Length 272;
  Best Local Similarity  22.9%;  Pred. No. 1.7e-05;
  Matches   62;  Conservative   44;  Mismatches  103;  Indels   62;  Gaps    8;

Qy          1 MILELDCGNS----------------LIKWRVIEGAARSVAGGLAESDDALV----EQLT 40
              |:| :|  |:                : :|| |   :   |  || : | |:    |:||
Db          1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQQWR-IRTESEVTADELALTIDGLIGEDSERLT 59

Qy         41 SQQALPVRACRLVSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYL----DYQ 96
                ||      |  ||           |:| :|    |        ||| |      : :
Db         60 GTAALSTVPSVLHEVR---------IMLDQYWPSVPHVLIE----PGVRTGIPLLVDNPK 106

Qy         97 RLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTH 156
              :| || :  :||:   :|| :|:| |:::  |:|:| |  ||| | ||: :
Db        107 EVGADRIVNCLAAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAAR 166

Qy        157 TRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLT 216
              :  :|  :    |::     |: | | :: | :    |                    |
Db        167 SAALRRVELARPRSVV----GKNTVECMQAGAVFGFAG-------------------LV 202

Qy        217 GGDAELVRDELAGARIMPDLVFVGLALACPI 247
                |    :|::::|  :  |:  |      |:
Db        203 DGLVGRIREDVSGFSVDHDVAIVATGHTAPL 233


RESULT 12
D72320
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C;Species: Thermotoga maritima
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: D72320
R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
Nature 399, 323-329, 1999
A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A;Reference number: A72200; MUID:99287316; PMID:10360571
A;Accession: D72320
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-246 <ARN>
A;Cross-references: GB:AE001754; GB:AE000512; NID:g4981417; PIDN:AAD35964.1; PID:g498141
A;Experimental source: strain MSB8
C;Genetics:
A;Gene: TM0883
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match           11.6%;  Score 145;  DB 2;  Length 246;
  Best Local Similarity  24.3%;  Pred. No. 3.4e-05;
  Matches   67;  Conservative   42;  Mismatches  103;  Indels   64;  Gaps   11;

Qy          1 MILELDCGN-----------SLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVR 48
              | | :| ||           :  :||       :: |: :::| |   |       :|
Db          1 MYLLVDVGNTHSVFSITEDGKTFRRWR--------LSTGVFQTEDELFSHLHPLLGDAMR 52

Qy         49 ACRLVSVRSEQETSQLVAR--LEQLFPVSALVASSGKQLAGVRNGYLDY-----QRLGLD 101
                : | |   |   |     :: | :| :   :      :|| : :       :| |
Db         53 EIKGIGVASVVPTQNTVIERFSQKYFHISPIWVKA-------KNGCVKWNVKNPSEVGAD 105

Qy        102 RWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIR 161
              |   :||      |  ::||:||| | |||  :| : || | ||  :|   |   |
Db        106 RVANVVAFVKEYGKNGIIIDMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGT---- 160

Qy        162 YDDAEARRALASLQP-----GQATAE----AVERGCLLMLRGF---VREQYAMACELLGP 209
                  |:  |  ::|     |: | |     |  | :  | |    ::| |
Db        161 -----AKLPLVEVKPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYG-------- 207

Qy        210 DCEIFLTGGDAELVRDELAGARIMPDLVFVGLALAC 245
              |  : |||| :::|:| :       ||   |:   |
Db        208 DLPVVLTGGQSKIVKDMIKHEIFDEDLTIKGVYHFC 243


RESULT 13
T36391
hypothetical protein SCE94.31c - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
C;Accession: T36391
R;Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, April 1999
A;Reference number: Z21573
A;Accession: T36391
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-265 <OLI>
A;Cross-references: EMBL:AL049628; PIDN:CAB40880.1; GSPDB:GN00070; SCOEDB:SCE94.31c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SCE94.31c
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match           11.1%;  Score 138.5;  DB 2;  Length 265;
  Best Local Similarity  23.8%;  Pred. No. 0.00014;
  Matches   70;  Conservative   42;  Mismatches   93;  Indels   89;  Gaps   13;

Qy          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60
              |:| :| ||:                || : :| :||               :|  | :
Db          1 MLLTIDVGNT------------HTVLGLFDGED-IVEHWR-----------ISTDSRRT 35

Qy         61 TSQLVARLEQLFPVSALVASSGKQLA---------------------------------- 86
                 :|   |: |  :  |:   | :|
Db         36 ADELAVLLQGLMGMHPLL---GDELGDGIDGIAICATVPSVLHELREVTRRYYGDVPAVL 92

Qy         87 ---GVRNG---YLDY-QRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLG 139
                 ||: |     |: : :| || :  |||  |     :|:| ||| | | |:| | ::|
Db         93 VEPGVKTGVPILTDHPKEVGADRIINAVAAVELYGGPAIVVDFGTATTFDAVSARGEYIG 152

Qy        140 GYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVRE- 198
              | | ||: :    |     ::|  :    |::     |: | ||:: |   :: ||  :
Db        153 GVIAPGIEISVEALGVKGAQLRKIEVARPRSVI----GKNTVEAMQSG---IVYGFAGQV 205

Qy        199 ---QYAMACELLG--PDCEIFLTGGDAELVRDELAGARIM----PDLVFVGLAL 243
                   || ||     |  :  |:| | :|   |  : ::    | |  :|| |
Db        206 DGVVNRMARELADDPDDVTVIATGGLAPMV---LGESSVIDEHEPWLTLMGLRL 256


RESULT 14
H86937
conserved hypothetical protein ML0232 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H86937
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Hol
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: H86937
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092576; PIDN:CAC29740.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML0232
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match           11.0%;  Score 137.5;  DB 2;  Length 274;
  Best Local Similarity  24.8%;  Pred. No. 0.00017;
  Matches   64;  Conservative   37;  Mismatches  108;  Indels   49;  Gaps    8;

Qy          1 MILELDCGNS----------------LIKWRVIEGAARSVAGGLAESDDALV----EQLT 40
```

```
              |:| :|  |:                 : :|| |   :   |  ||   | |:    |:|
Db          1 MLLAIDVRNTHTVVGLLSGSKEHAKVVQQWR-IRTESEVTADELALIIDGLIGDDSERLA 59

Qy         41 SQQALPVRACRLVSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYL----DYQ 96
               ||      |  ||           |:| :|    |        |||  |      : :
Db         60 GAAALSTVPSVLHEVR---------IMLDQYWPSVPHVLIE----PGVRTGIPLLVDNPK 106

Qy         97 RLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTH 156
              :| || :  :|| |   :| :|:| |:::  |:|:| |  ||| | ||: :
Db        107 EVGADRIVNCLAAFHKFGQAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAAR 166

Qy        157 TRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFV-------REQYAMACELLGP 209
              :  :|  :    |::    |: | | :: | :    | |       |:        ||
Db        167 SAALRRVELARPRSVV----GKNTVECMQAGVVFGFAGLVDGLVGRMRQDVEEFSGDLGN 222

Qy        210 DCEIFLTGGDAELVRDEL 227
              :  ||  | |:  ||
Db        223 RVAVVATGHTAPLLLPEL 240


RESULT 15
D71326
conserved hypothetical protein TP0431 - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 20-Jun-2000
C;Accession: D71326
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71326
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-273 <COL>
A;Cross-references: GB:AE001220; GB:AE000520; NID:g3322705; PIDN:AAC65417.1; PID:g332271
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0431
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match            10.9%;  Score 136;  DB 2;  Length 273;
  Best Local Similarity  28.4%;  Pred. No. 0.00023;
  Matches   56;  Conservative   27;  Mismatches   86;  Indels   28;  Gaps    6;

Qy         66 ARLEQLFPV-SALVASSGKQLAGVRN---GYLDYQRL-----------GLDRWLALVAA 109
              | :  : || :  :| :  |::||:    |   |: |           | |     |||
Db         63 AFISSVVPVLTKTIADAVAQISGVQPVVFGPWAYEHLPVRIPEPVRAEIGTDLVANAVAA 122

Qy        110 HHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARR 169
              :   : ||:|:| |||:|   |   |:  |  | ||:      | | | ::
Db        123 YVHFRSACVVVDCGTALTFTAVDGTGLIQGVAIAPGLRTAVQSLHTGTAQLPL----VPL 178

Qy        170 ALASLQPGQATAEAVE----RGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRD 225
              ||     |: |  ||:    || | ::|  :    |   : ||  |   :||| : |
Db        179 ALPDSVLGKDTTHAVQAGVVRGTLFVIRAMI----AQCQKELGCRCAAVITGGLSRLFSS 234

Qy        226 ELAGARIMPDLVFVGLA 242
              |:    |   |   |||
Db        235 EVDFPPIDAQLTLSGLA 251


Search completed: June 24, 2003, 22:03:17
Job time : 10.0456 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 4.31341 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-20
Perfect score: 1248
Sequence: 1 MILELDCGNSLIKWRVIEGA.....GARIMPDLVFVGLALACPIE 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	178.5	14.3	267	1	BAF_BORPE	Q45338 bordetella
2	111	8.9	233	1	YACB_BACSU	P37564 bacillus su
3	89.5	7.2	428	1	BCHN_RHOSH	Q9rfd4 rhodobacter
4	89	7.1	324	1	CYPK_STRVN	P16560 streptomyce
5	86.5	6.9	432	1	GSA_VIBCH	Q9ku97 vibrio chol
6	86.5	6.9	564	1	HUTU_TRIRP	P53385 trifolium r
7	85.5	6.9	611	1	GLMS_BRUME	Q8yc47 b glucosami
8	85	6.8	298	1	TRUB_MYCTU	O33335 mycobacteri
9	83.5	6.7	438	1	GSA_STRCO	Q9f2s0 streptomyce
10	83	6.7	238	1	YOXD_BACSU	P14802 bacillus su
11	83	6.7	766	1	METE_PSEAE	P57703 pseudomonas
12	82.5	6.6	251	1	TPIS_PSEAE	Q9hv51 pseudomonas
13	82.5	6.6	258	1	YL48_MYCTU	O06228 mycobacteri
14	82.5	6.6	588	1	HUTU_HALN1	Q9hqd8 halobacteri
15	82.5	6.6	624	1	SIR_SYNP7	P30008 synechococc
16	81	6.5	672	1	HYFB_ECOLI	P23482 escherichia
17	80.5	6.5	388	1	YC67_MYCTU	Q11052 mycobacteri
18	80.5	6.5	650	1	GSPD_ERWCA	P31701 erwinia car
19	80.5	6.5	875	1	YC78_MYCTU	Q11042 mycobacteri
20	80	6.4	419	1	PGK_TREPA	O83549 treponema p
21	80	6.4	576	1	CH62_MAIZE	Q43298 zea mays (m
22	80	6.4	577	1	CH61_MAIZE	P29185 zea mays (m
23	80	6.4	3591	1	FHAB_BORPE	P12255 bordetella
24	79.5	6.4	506	1	DHA2_ALCEU	P46368 alcaligenes
25	79.5	6.4	675	1	ATKB_DEIRA	Q9rzp0 deinococcus
26	79.5	6.4	923	1	GYRA_PSEAE	P48372 pseudomonas
27	79	6.3	3567	1	ERY2_SACER	Q03132 saccharopol
28	78.5	6.3	2194	1	GLSN_MEDSA	Q03460 medicago sa
29	78	6.2	259	1	MODD_RHOCA	Q08384 rhodobacter
30	78	6.2	265	1	THID_MYCTU	P96268 mycobacteri
31	78	6.2	370	1	HUPK_RHILV	P28153 rhizobium l
32	77.5	6.2	232	1	GPH_YERPE	Q8zjf3 yersinia pe
33	77.5	6.2	638	1	SRPR_HUMAN	P08240 homo sapien
34	77	6.2	393	1	HEMX_ECOLI	P09127 escherichia
35	77	6.2	577	1	CH60_ARATH	P29197 arabidopsis
36	77	6.2	672	1	Y959_MYCTU	P71551 mycobacteri
37	77	6.2	955	1	CHRD_HUMAN	Q9h2x0 homo sapien
38	77	6.2	1172	1	LMB3_HUMAN	Q13751 homo sapien
39	76.5	6.1	311	1	APE_MOUSE	P08226 mus musculu
40	76.5	6.1	369	1	RECF_PSEAE	Q9i7c3 pseudomonas
41	76.5	6.1	373	1	RECF_STRCO	P36176 streptomyce
42	76.5	6.1	426	1	GSA_SALTY	P21267 salmonella
43	76.5	6.1	3726	1	TRX_DROME	P20659 drosophila
44	76	6.1	208	1	HIS7_TRIHA	P34041 trichoderma
45	76	6.1	468	1	YOPH_YEREN	P15273 yersinia en

ALIGNMENTS

RESULT 1
BAF_BORPE
ID BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshazer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli.";
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN=BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a birA homolog in Bordetella pertussis.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN=BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus
RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis.";
RL Mol. Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC ----------
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ----------
DR EMBL; U12020; AAA75361.1; -.
DR EMBL; AF016461; AAC68834.1; -.
DR EMBL; X90711; CAA62242.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

```
  Query Match             14.3%;  Score 178.5;  DB 1;  Length 267;
  Best Local Similarity   29.6%;  Pred. No. 2.7e-08;
  Matches   75;  Conservative   32;  Mismatches   93;  Indels   53;  Gaps   11;

Qy          1 MILELDCGNSLIK--WRVIEG--AARSVAGGLAES--DDALVEQLTSQQALPVRACRLVS 54
              ||: :| ||| :|  |   :   |||  |    ::   |||   | :   || |  | :
Db          1 MIILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLAT---LPRRPQRALG 57

Qy         55 VRSEQETSQLVARLEQLFPVSALVASSGKQL---------AGVRNGYLDYQRLGLDRWLA 105
              |         || | :   ::| : : |  :           |:|||| :  :|| |||
Db         58 VN--------VAGLARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWAC 109

Qy        106 LVAAHHLAKK-----ACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRI 160
              :|    ||::       ||   ||| | | :  | |  || | ||  :|| |   |  :
Db        110 MVGV--LARQPSVHPPLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHL 167

Qy        161 RYDDA-------EARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLGPDCEI 213
                 |        :  :|:||   | | |:|          | :  |:    :  |   ||
Db        168 PLADGLVADYPIDTHQAIAS---GIAAAQA----------GAIVRQWLAGRQRYGQAPEI 214

Qy        214 FLTGGDAELVRDE 226
              :: ||    || |
Db        215 YVAGGGWPEVRQE 227


RESULT 2
YACB_BACSU
ID   YACB_BACSU     STANDARD;      PRT;   233 AA.
AC   P37564;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yacB.
GN   YACB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             8.9%;  Score 111;  DB 1;  Length 233;
  Best Local Similarity   21.0%;  Pred. No. 0.014;
  Matches   44;  Conservative   35;  Mismatches   79;  Indels   52;  Gaps    7;

Qy          1 MILELDCGNS-----------LIKWRV------IEGAARSVAGGLAESDDALVEQ---- 38
              |:| :| ||:              ||:       |     :   | :    : ||
Db          1 MLLVIDVGNTNTVLGVYHDGKLEYHWRIETSRHKTEDEFGMILRSLFDHSGLMFEQIDGI 60

Qy         39 LTSQQALPV-----RACRLVSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYL 93
              : |    |:     | |                   : | :   :   | :  |:   |
Db         61 IISSVVPPIMFALERMC-----------------TKYFHIEPQIVGPGMK-TGLNIKYD 101

Qy         94 DYQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQL 153
              : : :| || :  ||| ||     :|:| ||| |   :  : ::|| | ||:|:    |
Db        102 NPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGITISTEAL 161

Qy        154 RTHTRRI------RYDDAEARRALASLQPG 177
                :   ::      | |:   :  ::::| |
Db        162 YSRAAKLPRIEITRPDNIIGKNTVSAMQSG 191


RESULT 3
BCHN_RHOSH
ID   BCHN_RHOSH     STANDARD;      PRT;   428 AA.
AC   Q9RFD4; Q9Z5D8;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Light-independent protochlorophyllide reductase subunit N
DE   (EC 1.18.-.-) (LI-POR subunit N) (DPOR subunit N).
GN   BCHN.
OS   Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1063;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX   MEDLINE=20115911; PubMed=10648776;
RA   Choudhary M., Kaplan S.;
RT   "DNA sequence analysis of the photosynthesis region of Rhodobacter
RT   sphaeroides 2.4.1.";
RL   Nucleic Acids Res. 28:862-867(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
```

```
RA   Naylor G.W., Addlesee H.A., Gibson L.C.D., Hunter C.N.;
RT   "The photosynthesis gene cluster of Rhodobacter sphaeroides.";
RL   Photosyn. Res. 62:121-139(1999).
CC   -!- FUNCTION: Uses Mg-ATP and reduced ferredoxin to reduce ring D of
CC       protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide).
CC       The bchN-bchB pair binds Pchlide. This reaction is light-
CC       independent.
CC   -!- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
CC   -!- SUBUNIT: Protochlorophyllide reductase is thought to be composed
CC       of three subunits; bchL, bchN and bchB. Could form a
CC       heterotetramer of two bchB and two bchN subunits.
CC   -!- SIMILARITY: BELONGS TO THE BCHN / CHLN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF195122; AAF24275.1; -.
DR   EMBL; AJ010302; CAB38725.1; -.
DR   TIGRFAMs; TIGR01279; DPOR_bchN; 1.
KW   Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.
FT   CONFLICT     62     62       A -> G (IN REF. 2).
FT   CONFLICT    223    301       MISSING (IN REF. 2).
FT   CONFLICT    306    306       S -> R (IN REF. 2).
FT   CONFLICT    348    348       P -> A (IN REF. 2).
FT   CONFLICT    415    415       P -> A (IN REF. 2).
SQ   SEQUENCE   428 AA;  46118 MW;  B357968A7D3B60D5 CRC64;

  Query Match             7.2%;  Score 89.5;  DB 1;  Length 428;
  Best Local Similarity   22.9%;  Pred. No. 2;
  Matches   64;  Conservative   25;  Mismatches  108;  Indels   83;  Gaps   11;

Qy         15 RVIEGAARSVAGGLAESDDA-LVEQLTSQQALPVRACRLVSVRSEQETSQLVARLEQL-- 71
              ||   :  :    : :|| |   : :  |  |   ||    :    | |: | ||
Db        140 RVYNFSGSGIETTFTQGEDACLASIVPTLPATEARELLLVGALPDVVEDQAVSLLTQLGI 199

Qy         72 FPVSALVASSGKQLAGV--------------------RNGY------LDYQRLGLDRWL 104
              ||  | |    :  ||                    | |         :   |   ||
Db        200 GPVRCLPAHHAAEAPGVGPNTVFALVQPFLGDTHGALTRRGARHIAAPFPFGEEGTTLWL 259

Qy        105 ALVAAHHLAKKACLVIDLGTAVTS-------DLVAADGVHLGG---YICPGMTLMRSQLR 154
                :    | :  :  |   |||:         |||    | |   :  |
Db        260 KAI-----ADEFGVSADTFEAVTAAPRARARKAVAAASEGLRGKSVFFLP---------- 304

Qy        155 THTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLGPDCE-- 212
                  |::   :||     :    ||| |   : ||           :|||| :
Db        305 ---------DSQLEPSLARFLTRECGMSAVEVGTPFLHRG----------ILGPDLDLI 344

Qy        213 ----IFLTGGDAELVRDELAGARIMPDLVFVGLALACPIE 248
                  :   | | |   | :  ||  |||   || || |:|
Db        345 AEGPVLSEGQDVERQLDRVRAAR--PDLTVCGLGLANPLE 382


RESULT 4
CYPK_STRVN
ID   CYPK_STRVN     STANDARD;      PRT;   324 AA.
AC   P16560;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Granaticin polyketide synthase bifunctional cyclase/dehydratase.
OS   Streptomyces violaceoruber.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1935;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Tu22;
RX   MEDLINE=90060034; PubMed=2583128;
RA   Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J.,
RA   Hopwood D.A.;
RT   "Structure and deduced function of the granaticin-producing
RT   polyketide synthase gene cluster of Streptomyces violaceoruber
RT   Tu22.";
RL   EMBO J. 8:2717-2725(1989).
CC   -!- FUNCTION: IS NEEDED FOR CORRECT CYCLIZATION OF THE OLIGOKETIDE
CC       LEADING TO ISOCHROMANEQUINONE FORMATION.
CC   -!- PATHWAY: Polyketide antibiotic granaticin biosynthesis.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16300; CAA34372.1; -.
DR   EMBL; X16144; CAA34267.1; -.
DR   PIR; S05396; S05396.
DR   InterPro; IPR005031; Polyket_cyc.
DR   Pfam; PF03364; Polyketide_cyc; 2.
KW   Antibiotic biosynthesis.
SQ   SEQUENCE   324 AA;  35673 MW;  4893A828DD1CB075 CRC64;

  Query Match             7.1%;  Score 89;  DB 1;  Length 324;
  Best Local Similarity   26.4%;  Pred. No. 1.6;
  Matches   60;  Conservative   21;  Mismatches   84;  Indels   62;  Gaps   12;

Qy         28 LAESDDALVEQLTSQQALPVRACRLVSVRSEQETSQLVARLEQLFPVSALVASSGKQL-- 85
              :|: | | | ::  ::||   : |        |   | |  |   ||  || | |
Db        128 VADDDPAAVARI--EEALDRNSAR--------ELGALAALAETGHPVDELVFSFTDTLPL 177

Qy         86 -AGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTS----DLVAADGVHLGG 140
                 |: |   :|   :|| |:   |:|:  |     || :      | | |||
Db        178 QGAARDAYTFVER--AERWAELLP--HVAQ--CGADRAGTGLEQWLEMDTVTADG----- 226

Qy        141 YICPGMTLMRSQLRTHTRR----------IRYDDAEARRALASLQPGQATAEAVERGCLL 190
                        ||| |          | |::    | | |   |: |      | :
Db        227 ------------STHTTRSARICRAPEWIAYNEQRTPR-LVSGHSGEWTFAQTPEGPVA 272

Qy        191 MLRGFVREQYAMACELLGPDCEIFLTGGDAE------LVRDELAGAR 231
                |  |  :   |:||||     |  ||       | |: ||  |
Db        273 TARHTVAVDPSGITEVLGPDA----TLADARAYLRDALGRNSLATLR 315


RESULT 5
GSA_VIBCH
ID   GSA_VIBCH      STANDARD;      PRT;   432 AA.
AC   Q9KU97;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)
DE   (Glutamate-1-semialdehyde aminotransferase) (GSA-AT).
GN   HEML OR VC0626.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=El Tor N16961 / Serotype O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
```

```
RL   Nature 406:477-483(2000).
CC   -!- CATALYTIC ACTIVITY: (S)-4-amino-5-oxopentanoate = 5-
CC       aminolevulinate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; second step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE004148; AAF93792.1; -.
DR   HSSP; P24630; 2GSA.
DR   TIGR; VC0626; -.
DR   InterPro; IPR000954; Aminotran_3.
DR   InterPro; IPR004639; HemL.
DR   Pfam; PF00202; aminotran_3; 1.
DR   TIGRFAMs; TIGR00713; hemL; 1.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW   Porphyrin biosynthesis; Isomerase; Pyridoxal phosphate;
KW   Complete proteome.
FT   BINDING     265    265       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   432 AA;  46181 MW;  963BD21BE460C714 CRC64;

  Query Match             6.9%;  Score 86.5;  DB 1;  Length 432;
  Best Local Similarity   22.1%;  Pred. No. 3.7;
  Matches   56;  Conservative   37;  Mismatches   83;  Indels   77;  Gaps   12;

Qy          9 NSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQETSQLVARL 68
              :::|:  ||: | | :: |     :  : :|                      |:||  :
Db         68 HAVIREAVIQAAQRGLSFGAPTEMEITMAEL--------------------VSELVPSM 106

Qy         69 EQLFPVSALVASSGKQ--LAGVR--NGYLDYQRLGLDRWLALVAAHHLAKKACLVIDLGT 124
              |||      : :|| :  :: :|   ||      | |: :     :|    : ||
Db        107 EQL-----RMVNSGTEATMSAIRLARGY-----TGRDKIIKFEGCYHGHADSLLVKAGSG 156

Qy        125 AVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAV 184
              |:|          ||    ||:   :    ||   |::| :: | | :   |:     |
Db        157 ALT----------LGQPSSPGVP---ADFAKHTLTARFNDLDSVRELFAANQGEIACIIV 203

Qy        185 E-----RGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDE--------LAGAR 231
              |       |:  : ||  |     |:  |           | |: ||        | ||:
Db        204 EPVAGNMNCIPPVEGF-HEGLREICDQEG----------ALLIFDEVMTGFRVALGGAQ 251

Qy        232 ----IMPDLVFVG 240
                  | |||  :|
Db        252 AHYNIKPDLTTLG 264


RESULT 6
HUTU_TRIRP
ID   HUTU_TRIRP     STANDARD;      PRT;   564 AA.
AC   P53385;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Urocanate hydratase (EC 4.2.1.49) (Urocanase) (Imidazolonepropionate
DE   hydrolase).
OS   Trifolium repens (Creeping white clover).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Trifolium.
OX   NCBI_TaxID=3899;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Huia;
RX   MEDLINE=93011982; PubMed=1356832;
RA   Koberstaedt A., Lenz M., Retey J.;
RT   "Isolation, sequencing and expression in E. coli of the urocanase
RT   gene from white clover (Trifolium repens).";
RL   FEBS Lett. 311:206-208(1992).
CC   -!- CATALYTIC ACTIVITY: 4,5-dihydro-4-oxo-5-imidazolepropanoate =
CC       urocanate + H(2)O.
CC   -!- COFACTOR: Binds 1 NAD per subunit.
CC   -!- PATHWAY: Histidine degradation; second step.
CC   -!- SUBUNIT: Homodimer.
CC   -!- SIMILARITY: BELONGS TO THE UROCANASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68950; CAA48765.1; -.
DR   InterPro; IPR000193; Urocanase.
DR   Pfam; PF01175; Urocanase; 1.
DR   ProDom; PD025423; Urocanase; 1.
DR   TIGRFAMs; TIGR01228; hutU; 1.
DR   PROSITE; PS01233; UROCANASE; 1.
KW   Histidine metabolism; Lyase; NAD.
FT   ACT_SITE    416    416       BY SIMILARITY.
SQ   SEQUENCE   564 AA;  61402 MW;  5A5E0FAC91622DC3 CRC64;

  Query Match             6.9%;  Score 86.5;  DB 1;  Length 564;
  Best Local Similarity   22.3%;  Pred. No. 4.9;
  Matches   64;  Conservative   32;  Mismatches   98;  Indels   93;  Gaps   13;

Qy          4 ELDCGNSLIKWRVIEGAARSVAGGL----AESDDALVEQLTSQQALPVRACRLVSVRSEQ 59
              || | |    | :|| | | :   |    ||  : ||      :|    ||
Db         20 ELHCAN----W-LIEAAYRMIQNNLDPDVAERPEDLVVYGGIGKAARNWAC--------- 65

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGL------DRWLALVAAHHLA 113
                 |::  |: | |   |:  ||| : ||   : |  |: :        |      | |
Db         66 -FEQILRSLQALQPEETLLVQSGKPV-GVFRTHADAPRVLIANSNLVPHWATWDHFHELD 123

Qy        114 KKACLVIDLGTAVTSDLVAADGV---------------------------HLGGY---- 141
              |   ::    || :   : | |:                           |||
Db        124 KAGLMMYGQMTAGSWIYIGAQGIVQGTFETFVEAGRKHYNGDLTGKWILTAGLGGMGGAQ 183

Qy        142 ----ICPGMTLM-------RSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLL 190
                  :  |  ::        |   |  || : :       |||           ::: |
Db        184 PLAGVLAGACVLAVECQESRIDFRLRTRYLDHKAFSVDEALA----------IIDKAC-- 231

Qy        191 MLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLV 237
                    :|: |::  |||   ||       |||:   ||  : ||:|
Db        232 ------KEKRAISVGLLGNAAEIL-----PELVQRAKAGG-MKPDIV 266


RESULT 7
GLMS_BRUME
ID   GLMS_BRUME     STANDARD;      PRT;   611 AA.
AC   Q8YC47;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glucosamine--fructose-6-phosphate aminotransferase [isomerizing]
DE   (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
DE   phosphate amidotransferase) (GFAT) (L-glutamine-D-fructose-6-phosphate
DE   amidotransferase) (Glucosamine-6-phosphate synthase).
GN   GLMS OR BMEII0685.
OS   Brucella melitensis.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Brucellaceae; Brucella.
OX   NCBI_TaxID=29459;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=16M / ATCC 23456 / Biotype 1;
RX   MEDLINE=20020109; PubMed=11756688;
RA   DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujer C., Los T.,
RA   Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA   Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
RA   Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA   Haselkorn R., Kyrpides N., Overbeek R.;
RT   "The genome sequence of the facultative intracellular pathogen
RT   Brucella melitensis.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC   -!- FUNCTION: Catalyzes the first step in hexosamine metabolism,
CC       converting fructose-6P into glucosamine-6P using glutamine as a
CC       nitrogen source (By similarity).
CC   -!- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-
CC       glutamate + D-glucosamine 6-phosphate.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SIS FAMILY.
CC       GFAT SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 TYPE-2 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE009704; AAL53927.1; -.
DR   InterPro; IPR000583; GATase_2.
DR   InterPro; IPR001347; SIS.
DR   Pfam; PF00310; GATase_2; 1.
DR   Pfam; PF01380; SIS; 2.
DR   TIGRFAMs; TIGR01135; glmS; 1.
DR   PROSITE; PS00443; GATASE_TYPE_II; 1.
KW   Transferase; Aminotransferase; Glutamine amidotransferase;
KW   Complete proteome.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    244       GLUTAMINE AMIDOTRANSFERASE.
FT   ACT_SITE      6      6       GATASE (BY SIMILARITY).
FT   ACT_SITE    606    606       ISOMERIZATION FRU-6P (BY SIMILARITY).
SQ   SEQUENCE   611 AA;  66814 MW;  329BE30E7079D872 CRC64;

  Query Match             6.9%;  Score 85.5;  DB 1;  Length 611;
  Best Local Similarity   25.4%;  Pred. No. 6.6;
  Matches   61;  Conservative   27;  Mismatches   63;  Indels   89;  Gaps   16;

Qy         50 CRLVSVRSEQETS-QLVARLEQLFPVSALVASSGKQLAGV---RNGYLDYQR-------- 97
              | :: :    | : :||  |::|          |   ||:   :|| || :|
Db          6 CGIIGIIGNDEVAPRLVDALKRL-------EYRGYDSAGIATLQNGRLDRRRAEGKLVNL 58

Qy         98 -------------LGLDRWLALVAAH-----HLAKKACLVIDLGTAVTSDLVAADGVHL 138
                         :|  ||    | |     | |           :|: |     ||
Db         59 EKRLAGEPLPGVIGIGHTRW----ATHGRPVEHNAH---------PHITTRLAV---VHN 102

Qy        139 GGYICPGMTLMRSQLRTHTRRIRYD---DAEARRALASLQPGQATAEAVERGCLLMLRGF 195
               |  |      :|: |    |:   :  :| |      |: |::  ||| | ||  |:|
Db        103 G--IIENFAELRAMLEAEGRKFETETDTEAVAHLVTRELEKGKSPVEAV-RDCLPHLKG- 158

Qy        196 VREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDL--------VFVG---LALA 244
                 :|:|          ||  || ||    | |||  | |         :|:|   :|||
Db        159 ---AFALA----------FLFEGDEEL----LIGARQGPPLAVGYGEGEMFLGSDAIALA 201


RESULT 8
TRUB_MYCTU
ID   TRUB_MYCTU     STANDARD;      PRT;   298 AA.
AC   O33335;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE   synthase) (Psi55 synthase) (Pseudouridylate synthase) (Uracil
DE   hydrolyase).
GN   TRUB OR RV2793C OR MT2862.1 OR MTV002.58C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC       PSI GC LOOP OF TRANSFER RNAS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC       5'-phosphate + H(2)O.
CC   -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL008967; CAA15588.1; -.
DR   EMBL; AE007112; AAK47182.1; -.
DR   TIGR; MT2862.1; -.
DR   TubercuList; Rv2793c; -.
DR   InterPro; IPR004510; TruB.
DR   InterPro; IPR002501; TruB_N.
DR   Pfam; PF01509; TruB_N; 1.
DR   TIGRFAMs; TIGR00431; TruB; 1.
KW   Lyase; tRNA processing; Complete proteome.
SQ   SEQUENCE   298 AA;  31819 MW;  9F77797DC13B34C6 CRC64;

  Query Match             6.8%;  Score 85;  DB 1;  Length 298;
  Best Local Similarity   22.7%;  Pred. No. 3.2;
  Matches   56;  Conservative   40;  Mismatches   93;  Indels   58;  Gaps   14;

Qy         16 VIEGAARSVAGGLAESDDALVEQLTSQ--QALPVRACRLVSVRSEQETSQLVARLEQL-F 72
              ::  | :| |  :         |    |  |::| :   : :: :  |  :|   : |:
Db         62 LLTAAPKSYAATIRLGQTTSTEDAEGQVLQSVPAKHLTIEAIDAAME--RLRGEIRQVPS 119

Qy         73 PVSALVASSGKQLAGVRNG---YLDYQRLGLDRWLALVAAHHLAKKACLVIDL------- 122
               |||:     :     | |    |: : : :||:  |      |::   :||:
Db        120 SVSAIKVGGRRAYRLARQGRSVQLEARPIRIDRFELLA-----ARRRDQLIDIDVEIDCS 174

Qy        123 -GT---AVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQ 178
               ||   |:  ||   | : :||::    | :|   ||   |   | |   |:|  |
Db        175 SGTYIRALARDL--GDALGVGGHV----TALR---RTRVGRFELDQA---RSLDDLAERP 222
```

```
Qy        179 ATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVF 238
              | : :::  ||||   | |                ||  :|       |  | :| :
Db        223 ALSLSLDEACLLM---FARRD--------------LTAAEASAA----ANGRSLPAVGI 260

Qy        239 VGLALAC 245
               |:  ||
Db        261 DGVYAAC 267


RESULT 9
GSA_STRCO
ID   GSA_STRCO      STANDARD;      PRT;   438 AA.
AC   Q9F2S0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Glutamate-1-semialdehyde 2,1-aminomutase (EC 5.4.3.8) (GSA)
DE   (Glutamate-1-semialdehyde aminotransferase) (GSA-AT).
GN   HEML OR SCO4469 OR SCD65.12.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
CC   -!- CATALYTIC ACTIVITY: (S)-4-amino-5-oxopentanoate = 5-
CC       aminolevulinate.
CC   -!- COFACTOR: PYRIDOXAL PHOSPHATE.
CC   -!- PATHWAY: Porphyrin biosynthesis by the C5 pathway; second step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: BELONGS TO CLASS-III OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC       AMINOTRANSFERASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL392176; CAC08376.1; -.
DR   HSSP; P24630; 2GSA.
DR   InterPro; IPR000954; Aminotran_3.
DR   InterPro; IPR004639; HemL.
DR   Pfam; PF00202; aminotran_3; 1.
DR   TIGRFAMs; TIGR00713; hemL; 1.
DR   PROSITE; PS00600; AA_TRANSFER_CLASS_3; 1.
KW   Porphyrin biosynthesis; Isomerase; Pyridoxal phosphate;
KW   Complete proteome.
FT   BINDING     274    274       PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ   SEQUENCE   438 AA;  45341 MW;  2B7A4909C4BD1197 CRC64;

  Query Match             6.7%;  Score 83.5;  DB 1;  Length 438;
  Best Local Similarity   23.4%;  Pred. No. 6.7;
  Matches   55;  Conservative   29;  Mismatches   96;  Indels   55;  Gaps    9;

Qy         16 VIEGAARSVAGGL-AESDDALVEQLTSQQALPVRACRLVSVRSEQETSQLVARLEQLFPV 74
                 | |  ||  : |  || : || | :  : | |:   ||||  |  | :    || : |
Db         84 VQEAVARGTSFGTPAEGEVALAEAMVERVA-PLEEVRLVS--SGTEATMSAIRLARGFTR 140

Qy         75 SALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAAD 134
                    |     : ||  :|::|                                  |:||
Db        141 RTKVI----KFAGCYHGHVD---------------------------------SLLAAA 162

Qy        135 GVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVE-----RGCL 189
              |  :  :  |    :     : |  : |:| :|  |  :  ||:      |      | :
Db        163 GSGVATFALPDTPGVTGAQASDTIVLPYNDLDAVHAAFAAHPGEIACVITEASPGNMGVV 222

Qy        190 LMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAG----ARIMPDLVFVG 240
                | ||  :    ||   |    :|::       |  |||      ::|||:  |
Db        223 PPLPGF-NQGLKDACAANG---ALFVSDEVMTGFRTSLAGWFGVEGVVPDLMTFG 273


RESULT 10
YOXD_BACSU
ID   YOXD_BACSU     STANDARD;      PRT;   238 AA.
AC   P14802;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical oxidoreductase yoxD (EC 1.-.-.-).
GN   YOXD.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=91192601; PubMed=1849493;
RA   Ahn K.S., Wake R.G.;
RT   "Variations and coding features of the sequence spanning the
RT   replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL   Gene 98:107-112(1991).
RN   [2]
RP   SEQUENCE OF 62-238 FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=88040469; PubMed=3118336;
RA   Carrigan C.M., Haarsma J.A., Smith M.T., Wake R.G.;
RT   "Sequence features of the replication terminus of the Bacillus
RT   subtilis chromosome.";
RL   Nucleic Acids Res. 15:8501-8509(1987).
CC   -!- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC       (SDR) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06168; CAA29533.1; -.
DR   EMBL; Z99114; CAB13743.1; -.
DR   PIR; S01270; S01270.
DR   HSSP; O70351; 1E3S.
DR   SubtiList; BG11048; yoxD.
DR   InterPro; IPR002198; ADH_short.
DR   Pfam; PF00106; adh_short; 1.
DR   PRINTS; PR00080; SDRFAMILY.
DR   PROSITE; PS00061; ADH_SHORT; 1.
KW   Hypothetical protein; Oxidoreductase; Complete proteome.
FT   NP_BIND      10     34       NAD OR NADP (BY SIMILARITY).
FT   ACT_SITE    155    155       BY SIMILARITY.
SQ   SEQUENCE   238 AA;  25299 MW;  9CC7A8D1204DF248 CRC64;

  Query Match             6.7%;  Score 83;  DB 1;  Length 238;
  Best Local Similarity   24.2%;  Pred. No. 3.7;
  Matches   45;  Conservative   32;  Mismatches   59;  Indels   50;  Gaps   11;
```

```
Qy          13 KWRVIEGAARSVAG------------GLAESDDALVEQLTSQ-QALPVRAC-RLVSVRS 57
Db           7 KTALITGGGRGIGRATALALAKEGVNIGLIGRTSANVEKVAEEVKALGVKAAFAAADVKD 66

Qy          58 EQETSQLVARL-EQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRW-----LALVAAHH 111
Db          67 ADQVNQAVAQVKEQLGDIDILINNAGISKFG---GFLD---LSADEWENIIQVNLMGVYH 120

Qy         112 ---------LAKKACLVIDL-------GTAVTSDLVAADGVHLGGYICPGMT-LMRSQLR 154
Db         121 VTRAVLPEMIERKAGDIINISSTAGQRGAAVTSAYSASK------FAVLGLTESLMQEVR 174

Qy         155 THTRRI 160
Db         175 KHNIRV 180
```

RESULT 11
METE_PSEAE
ID METE_PSEAE STANDARD; PRT; 766 AA.
AC P57703; Q9I2H7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase
DE (EC 2.1.1.14) (Methionine synthase, vitamin-B12 independent isozyme)
DE (Cobalamin-independent methionine synthase).
GN METE OR PA1927.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: CATALYZES THE TRANSFER OF A METHYL GROUP FROM 5-
CC METHYLTETRAHYDROFOLATE TO HOMOCYSTEINE RESULTING IN METHIONINE
CC FORMATION (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 5-methyltetrahydropteroyltri-L-glutamate + L-
CC homocysteine = tetrahydropteroyltri-L-glutamate + L-methionine.
CC -!- COFACTOR: ZINC; BINDS ONE ION PER SUBUNIT (BY SIMILARITY).
CC -!- PATHWAY: TERMINAL STEP IN THE DE NOVO BIOSYNTHESIS OF METHIONINE.
CC -!- SIMILARITY: BELONGS TO THE VITAMIN-B12 INDEPENDENT METHIONINE
CC SYNTHASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AE004619; AAG05315.1; -.
DR InterPro; IPR002629; Methionine_synt.
DR Pfam; PF01717; Methionine_synt; 1.
KW Transferase; Methyltransferase; Methionine biosynthesis; Zinc; Repeat;
KW Complete proteome.
FT METAL 650 650 ZINC (BY SIMILARITY).
FT METAL 652 652 ZINC (BY SIMILARITY).
FT METAL 735 735 ZINC (BY SIMILARITY).
SQ SEQUENCE 766 AA; 86210 MW; 7D638917B844AE81 CRC64;

Query Match 6.7%; Score 83; DB 1; Length 766;
Best Local Similarity 22.0%; Pred. No. 14;
Matches 62; Conservative 38; Mismatches 90; Indels 92; Gaps 10;

```
Qy          30 ESDDALVEQLTSQQALPVRACRLVSV----RSEQETSQLVARLEQLFPVSALVASSGK-- 83
Db         338 EREDGLDAELKSWLAFAVQKCREVAVLARAATEPEAAEVLAALEESRAVQASRASSPRIH 397

Qy          84 ------QLAGVRNG----------YLDYQRLGLD-------------RWLALVAAHHLAK 114
Db         398 KPAVQARLAAIKASDAQRRSPFAERIARQRAGLDLPAFPTTTIGSFPQTSSIRLARQSFK 457

Qy         115 KACL-------------------------VIDLGTAVTSDLVAADGVHLGGYICPGMTL 148
Db         458 QGKLSEAEYIEAMHSEIRHAVQIQEQLGLDVLVHGEAERNDMVEYFAEQLDGYVFTRFGW 517

Qy         149 MRSQ-LRTHTRRIRYDDAEARRALAS--LQPGQATAEAVERGCL-----LMLRGFVRE-- 198
Db         518 VQSYGSRCVKPAVIYGDLSRPRAMTVEWIRYAQSLTDKVMKGMLTGPVTMLMWSFPREDV 577

Qy         199 -------QYAMACELLGPDCEIFLTGGDAELVRDELAGARIM 233
Db         578 SREVQARQLALAIR-------------DEVVDLEAAGIRIV 605
```

RESULT 12
TPIS_PSEAE
ID TPIS_PSEAE STANDARD; PRT; 251 AA.
AC Q9HV51;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Triosephosphate isomerase (EC 5.3.1.1) (TIM).
GN TPIA OR PA4748.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC phosphate.
CC -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; AE004888; AAG08134.1; -.
DR HSSP; P04790; 1TRE.
DR InterPro; IPR000652; Triophos_ismrse.
DR Pfam; PF00121; TIM; 1.
DR ProDom; PD001005; Triophos_ismrse; 1.
DR TIGRFAMs; TIGR00419; tim; 1.
DR PROSITE; PS00171; TIM; 1.
KW Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;

```
KW   Pentose shunt; Complete proteome.
FT   ACT_SITE     95     95       BY SIMILARITY.
FT   ACT_SITE    167    167       BY SIMILARITY.
SQ   SEQUENCE   251 AA;  25621 MW;  8D988EAF1B35FA38 CRC64;

  Query Match             6.6%;  Score 82.5;  DB 1;  Length 251;
  Best Local Similarity   25.8%;  Pred. No. 4.4;
  Matches   54;  Conservative   23;  Mismatches   53;  Indels   79;  Gaps   12;

Qy         28 LAESDDALVEQLTSQQALPVRACRLVSV------RSEQE---TSQLVARLEQLFPVSALV 78
              | |||: :  :  : |     :| || |      |:|:|   | ::|||
Db        102 LGESDEVVSRKFAAAQ-----SCGLVPVLCVGETRAEREAGKTLEVVAR---------- 145

Qy         79 ASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHL 138
                 ||  |         ||:  :   | |:        |  :|| :|:    |  ||
Db        146 -----QLGSV------IDELGVGAFARAVVAYE------PVWAIGTGLTASPAQAQEVH- 187

Qy        139 GGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVRE 198
                    :|:||         ::||                  | :|  |:  | |:
Db        188 --------AAIRAQLAA-------ENAE----------------VAKGVRLLYGGSVKA 215

Qy        199 QYAMACELLG-PDCEIFLTGGDAELVRDE 226
               | | || | || :  | || | |  ||
Db        216 --ASAAELFGMPDIDGGLVGG-ASLNADE 241


RESULT 13
YL48_MYCTU
ID   YL48_MYCTU     STANDARD;      PRT;   258 AA.
AC   O06228;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein Rv2148c.
GN   RV2148C OR MT2207 OR MTCY270.20.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE UPF0001 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z95388; CAB08645.1; -.
DR   EMBL; AE007068; AAK46491.1; ALT_INIT.
DR   TIGR; MT2207; -.
DR   TubercuList; Rv2148c; -.
DR   InterPro; IPR001608; UPF0001.
DR   Pfam; PF01168; UPF0001; 1.
DR   TIGRFAMs; TIGR00044; UPF0001; 1.
DR   PROSITE; PS01211; UPF0001; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   258 AA;  27693 MW;  6C534D96E9368BB2 CRC64;

  Query Match             6.6%;  Score 82.5;  DB 1;  Length 258;
  Best Local Similarity   25.2%;  Pred. No. 4.5;
  Matches   52;  Conservative   25;  Mismatches   78;  Indels   51;  Gaps   10;

Qy         20 AARSVAGGLAESDDALVEQLTSQQALPVR-------------ACRLVSVRSEQETSQLV 65
              |  ::   || : :|    :   : ||:                || |    ||| |  :
Db         18 ALAAMRSRLAAAAEAAGRNVGEIELLPITKFFPATDVAILFRLGCRSVGESREQEASAKM 77

Qy         66 ARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLG-LDRWLALVAAH--HLAKKACLVIDL 122
              | | :|   : |  | |      |  | :   : | | ||     ||  |    : ||
Db         78 AELNRLLAAAELGHSGGVHWHMV--GRIQRNKAGSLARW-----AHTAHSVDSSRLV--- 127

Qy        123 GTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQATAE 182
               ||:   :|||   |  |          :|| :  ::  |   :|  : |  ||
Db        128 -TALDRAVVAALAEHRRG----------ERLRVYV-QVSLDGDGSRGGVDSTTPG----- 170

Qy        183 AVERGCLLMLRGFVREQYAMACELLG 208
              ||:| |       : | :  ||:|
Db        171 AVDRIC-------AQVQESEGLELVG 189


RESULT 14
HUTU_HALN1
ID   HUTU_HALN1     STANDARD;      PRT;   588 AA.
AC   Q9HQD8;
DT   15-JUN-2002 (Rel. 41, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Probable urocanate hydratase (EC 4.2.1.49) (Urocanase)
DE   (Imidazolonepropionate hydrolase).
GN   HUTU OR VNG1208G.
OS   Halobacterium sp. (strain NRC-1).
OC   Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC   Halobacteriaceae; Halobacterium.
OX   NCBI_TaxID=64091;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20504483; PubMed=11016950;
RA   Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA   Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA   Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA   Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA   Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA   Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA   Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA   Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RT   "Genome sequence of Halobacterium species NRC-1.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC   -!- CATALYTIC ACTIVITY: 4,5-dihydro-4-oxo-5-imidazolepropanoate =
CC       urocanate + H(2)O.
CC   -!- COFACTOR: Binds 1 NAD per subunit (By similarity).
CC   -!- PATHWAY: Histidine degradation; second step.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE UROCANASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE005047; AAG19577.1; -.
DR   InterPro; IPR000193; Urocanase.
DR   Pfam; PF01175; Urocanase; 1.
DR   ProDom; PD025423; Urocanase; 1.
DR   TIGRFAMs; TIGR01228; hutU; 1.
DR   PROSITE; PS01233; UROCANASE; 1.
KW   Histidine metabolism; Lyase; NAD; Complete proteome.
FT   ACT_SITE.   431    431       BY SIMILARITY.
SQ   SEQUENCE   588 AA;  63222 MW;  FD14A8F699D6F8C9 CRC64;

  Query Match             6.6%;  Score 82.5;  DB 1;  Length 588;
  Best Local Similarity   23.2%;  Pred. No. 11;
  Matches   57;  Conservative   27;  Mismatches   87;  Indels   75;  Gaps   12;

Qy          2 ILELDC-GNSLIKWRVIEG-------AARSVAGGLAESDDA-----LVEQLTSQQALPVR 48
              |  | | |    :|  : |         |:|     | ||      | ::    | || |
Db        370 IRPLFCRGKGPFRWVALSGNPADIHRTDRAVTELFPEKDDLHRWIDLAQEHVQFQGLPSR 429

Qy         49 ACRLVSVRSEQ---ETSQLVARLEQLF-------PV--------SALVASSGKQLAGVRN 90
               | |    ::    | ::   |: :|       |:        : |||  ::   :|:
Db        430 VCWLGYCAADDDLTERARFAVRINELVDNGEIEAPIVVTRDHLDAGSVASPNRETEAMRD 489

Qy         91 GYLDYQRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMR 150
              |       :  |  | |             | ||  :|:|:   || || :  | :|
Db        490 G-----TDAVADWPILNAL-----------LNTAAGADIVS---VHNGGGVGIGNSL-- 527

Qy        151 SQLRTHTRRIRYDD-----AEARRALASLQPG-----------QATAEAVERGCLLMLR 193
                  ||      |     ||  | : :  ||             |  ||   |  : :|
Db        528 -----HTNNHVVLDGSDAAAETARRVFTTDPGMGVIRHADAGYADALVEADASGVTVPMR 582

Qy        194 GFVREQ 199
                 |||
Db        583 DAEREQ 588


RESULT 15
SIR_SYNP7
ID   SIR_SYNP7      STANDARD;      PRT;   624 AA.
AC   P30008;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Sulfite reductase (Ferredoxin) (EC 1.8.7.1).
GN   SIR.
OS   Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX   NCBI_TaxID=1140;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93349955; PubMed=8347657;
RA   Gisselmann G., Klausmeier P., Schwenn J.D.;
RT   "The ferredoxin:sulphite reductase gene from Synechococcus PCC7942.";
RL   Biochim. Biophys. Acta 1144:102-106(1993).
CC   -!- CATALYTIC ACTIVITY: H(2)S + 3 oxidized ferredoxin + 3 H(2)O =
CC       sulfite + 3 reduced ferredoxin.
CC   -!- COFACTOR: THIS ENZYME CONTAINS ONE SIROHEME AND ONE 4FE-4S
CC       IRON-SULFUR CENTER AS PROSTHETIC GROUPS.
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: THE C-TERMINAL DOMAIN IS A 4FE-4S/SIROHEME DOMAIN
CC       FOUND IN NITRITE REDUCTASES (EC 1.6.6.4 AND EC 1.7.7.1) AND
CC       SULFITE REDUCTASES (EC 1.8.1.2 AND EC 1.8.7.1).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z11755; CAA77809.1; -.
DR   PIR; S19860; RDYCS7.
DR   HSSP; P17846; 4AOP.
DR   InterPro; IPR000660; Nir_Sir.
DR   InterPro; IPR005117; Nir_sir_fer.
DR   Pfam; PF01077; NIR_SIR; 1.
DR   Pfam; PF03460; NIR_SIR_ferr; 2.
DR   PRINTS; PR00397; SIROHAEM.
DR   PROSITE; PS00365; NIR_SIR; 1.
KW   Oxidoreductase; Heme; Iron-sulfur; 4Fe-4S.
FT   METAL       446    446       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       452    452       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       491    491       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       495    495       IRON-SULFUR (4FE-4S) AND SIROHEME
FT                                (BY SIMILARITY).
SQ   SEQUENCE   624 AA;  70031 MW;  A4AABAFD6BA2F3E2 CRC64;

  Query Match             6.6%;  Score 82.5;  DB 1;  Length 624;
  Best Local Similarity   22.9%;  Pred. No. 12;
  Matches   64;  Conservative   34;  Mismatches   90;  Indels   91;  Gaps   14;

Qy         17 IEGAARSVAGGLA------ESDDALVEQLTSQQALPVRACRLVSVRSEQE----TSQLVA 66
              :||     | ||:         |:   | : |    |  : |      | ::::    :::  |
Db        256 LEGFNVYVGGGMGRTHNKEETFARLADPLGFVPAADIYAAVQAIVATQRDYGDRSNRRHA 315

Qy         67 RLEQLFP-------VSALVASSGKQLAGVR--------NGYLDYQRLGLDRWLALVAAHHL 112
              |:: |           |: :  || :| ||          || :   |  :|
Db        316 RMKYLIHDWGIAKFKEAVESVFGKAIAPVRELPPFRYRDYLGWHEQGDGKWF-------- 367

Qy        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRR------------- 159
                     ||  :||  :  |             |   :|| ||    |
Db        368 ---------LGLPITSGRIKDD----------GNWQLRSALREIVSRWQLPLLLTGSQDV 408

Qy        160 IRYDDAEARRALASLQPGQATAEAVERGCLLMLRGF-------VREQYAMACELLGPDCE 212
              : ||          :|||     |:::  ||  ||          ::|||||  | | |
Db        409 LIYD----------VQPGDRA--AIDK--LLRDRGVHTVEAIDSLQRYAMACPAL-PTCG 453

Qy        213 IFLTGGDAE----LVR-DELAGARIMPDLVFVGLALACP 246
              : :|  :       |||   |   : :||  ||     ||
Db        454 LAITESERALPGLLVRIRRLLEEQGLPDEHFVVRMTGCP 492


Search completed: June 24, 2003, 21:51:27
Job time : 5.31341 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 17.2942 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-43
Perfect score: 3078
Sequence: 1 MTVLKLSHWRVLAELADGLP.....LVIYGLLNMIAAEGREYEHI 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	194	6.3	311	4	US-09-175-172-2	Sequence 2, Appli
2	190.5	6.2	323	4	US-09-237-119-2	Sequence 2, Appli
3	158	5.1	328	4	US-09-134-001C-3229	Sequence 3229, Ap
4	116	3.8	5087	4	US-09-144-085-1	Sequence 1, Appli
5	114.5	3.7	3816	4	US-09-428-517-3	Sequence 3, Appli
6	111.5	3.6	6095	4	US-09-144-085-2	Sequence 2, Appli
7	109	3.5	5588	4	US-09-036-987A-6	Sequence 6, Appli
8	109	3.5	5588	4	US-09-370-700-6	Sequence 6, Appli
9	108.5	3.5	3170	2	US-07-642-734C-5	Sequence 5, Appli
10	108.5	3.5	3170	3	US-08-439-009A-5	Sequence 5, Appli
11	107	3.5	444	6	5310667-5	Patent No. 5310667
12	105.5	3.4	4613	4	US-09-105-537-31	Sequence 31, Appl
13	105.5	3.4	11877	4	US-09-105-537-6	Sequence 6, Appli
14	105	3.4	4928	4	US-09-036-987A-5	Sequence 5, Appli
15	105	3.4	4928	4	US-09-370-700-5	Sequence 5, Appli
16	103.5	3.4	1891	2	US-08-804-227C-12	Sequence 12, Appl
17	103.5	3.4	1891	2	US-08-804-198-6	Sequence 6, Appli
18	101.5	3.3	357	4	US-09-605-785-826	Sequence 826, App
19	101.5	3.3	1228	4	US-09-605-785-537	Sequence 537, App
20	101.5	3.3	1228	4	US-09-439-313-537	Sequence 537, App
21	101.5	3.3	1261	4	US-09-605-785-538	Sequence 538, App
22	101.5	3.3	1261	4	US-09-439-313-538	Sequence 538, App
23	101.5	3.3	5215	4	US-09-105-537-2	Sequence 2, Appli
24	101	3.3	1222	2	US-08-682-517-15	Sequence 15, Appl
25	101	3.3	1252	2	US-08-682-517-9	Sequence 9, Appli
26	99.5	3.2	444	1	US-08-476-008-56	Sequence 56, Appl
27	99.5	3.2	444	1	US-08-306-063-56	Sequence 56, Appl
28	99.5	3.2	444	1	US-08-833-485-56	Sequence 56, Appl
29	99.5	3.2	444	4	US-09-243-374-5	Sequence 5, Appli
30	99.5	3.2	444	4	US-09-137-440-56	Sequence 56, Appl
31	99.5	3.2	444	4	US-09-000-062-3	Sequence 3, Appli
32	99.5	3.2	444	4	US-09-000-062-5	Sequence 5, Appli
33	99.5	3.2	570	3	US-08-832-078-5	Sequence 5, Appli
34	99	3.2	4150	4	US-09-428-517-2	Sequence 2, Appli
35	98.5	3.2	4551	3	US-09-320-878-1	Sequence 1, Appli
36	95	3.1	1489	6	5183745-2	Patent No. 5183745
37	94	3.1	1525	4	US-09-396-651B-1	Sequence 1, Appli
38	94	3.1	3724	2	US-08-804-227C-10	Sequence 10, Appl
39	94	3.1	3724	2	US-08-804-198-4	Sequence 4, Appli
40	94	3.1	4472	2	US-08-804-227C-2	Sequence 2, Appli
41	94	3.1	4545	2	US-08-804-227C-14	Sequence 14, Appl
42	93.5	3.0	591	4	US-09-199-290-7	Sequence 7, Appli
43	93	3.0	1611	2	US-08-804-227C-5	Sequence 5, Appli
44	92.5	3.0	618	4	US-09-199-290-34	Sequence 34, Appl
45	92.5	3.0	2475	4	US-09-413-814-48	Sequence 48, Appl

ALIGNMENTS

RESULT 1
US-09-175-172-2
; Sequence 2, Application US/09175172
; Patent No. 6255075
; GENERAL INFORMATION:
; APPLICANT: Huang, Jianzhong
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: birA
; FILE REFERENCE: GM10172
; CURRENT APPLICATION NUMBER: US/09/175,172
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-175-172-2

Query Match 6.3%; Score 194; DB 4; Length 311;
Best Local Similarity 28.7%; Pred. No. 4.8e-13;
Matches 77; Conservative 46; Mismatches 99; Indels 46; Gaps 12;

```
Qy         76 LGERSGFQTALKHECASSNDEILELARIAPDKAHK--TICVTHLQSKGRGRQGRKWSHRL 133
                 | |    : : | | |:  :    |:.| |  |:  |: :   |: ||||  | :
Db         69 LEENLPIKVSFKPETKSTQLD----AKEAIDLGHEANTLYLASYQTAGRGRFQRSFYSPQ 124

Qy        134 GECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-DVQIKWPNDLVVGRDKLGGI 192
                 |   | :     : |  :| | : : | |  :|:  | | || ||| ||: :   |:|||
Db        125 GGIYM-TLHLKPNLPYDKLPSYTLLVAGAVYKAIKNLTLIDVDIKWVNDIYLNNHKIGGI 183

Qy        193 LIE---TVRTGGKTVAVVGIGINFVL---PKEVENAASVQSLFQTASRRGNADAAVLLET 246
                 | |   :| ||  |  ::|:|||| :   |:|::  |:  |||:       | | :
Db        184 LTEAMTSVETGLVTDIIIGVGINFTIKDFPQELKEKAA--SLFK-------ATAPITRNE 234

Qy        247 LLVEL---------DAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKG 297
                 |::|:         : :|  | :  |              || |    : :  ::|  |
Db        235 LIIEIWRAFFETPAEELLYLYKKQSFI-----------LGKEVTFTLE-QKDYKGLAKD 281

Qy        298 VDGQGVLHLETAEGKQT-VVSGEISLRS 324
                 :   | | ::   ||:  : ||||| |
Db        282 ISENGKLLVQCDNGKEIWLNSGEISLNS 309
```

RESULT 2
US-09-237-119-2
; Sequence 2, Application US/09237119
; Patent No. 6448038
; GENERAL INFORMATION:
; APPLICANT: Kallender, Howard

```
;  APPLICANT: Sylvester, Daniel R.
;  APPLICANT: McDevitt, Damien
;  APPLICANT: Huang, Jianzhong
;  APPLICANT: Van Horn, Stephanie
;  TITLE OF INVENTION: birA
;  FILE REFERENCE: GM10189
;  CURRENT APPLICATION NUMBER: US/09/237,119
;  CURRENT FILING DATE: 1999-01-26
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-237-119-2

  Query Match             6.2%;  Score 190.5;  DB 4;  Length 323;
  Best Local Similarity   26.7%;  Pred. No. 1.3e-12;
  Matches   80;  Conservative   43;  Mismatches  124;  Indels   53;  Gaps   14;

Qy         35 QQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSN 94
Db         61 QQLPDIWYQ------GIIDQY------TKSSALFDFSEVYDSIDST--QLAAKKSLVGNQ 106

Qy         95 DEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGS 154
Db        107 SSFF----ILSDE----------QTKGRGRFNRHWSSSKGQGLWMS---VVLRPNVAFSM 149

Qy        155 LSP---VAAVACRRALSRLGLD-VQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV-GI 209
Db        150 ISKFNLFIALGIRDAIQHFSQDEVKVKWPNDIYIDNGKVCGFLTEMVANNDGIEAIICGI 209

Qy        210 GINFVLPKEVEN--------AASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARD 261
Db        210 GIN--LTQQLENFDESIRHRATSIQ-----LHDKNKLDRYQFLERLLQEIEKRYNQFLTL 262

Qy        262 GFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK-QTVVSGEI 320
Db        263 PFSEIREEYIAASNIWNR-TLLFTENDKQFKGQAIDLDYDGYLIVRDEAGESHRLISADI 321


RESULT 3
US-09-134-001C-3229
; Sequence 3229, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3229
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3229

  Query Match             5.1%;  Score 158;  DB 4;  Length 328;
  Best Local Similarity   25.2%;  Pred. No. 6.1e-09;
  Matches   75;  Conservative   45;  Mismatches  140;  Indels   38;  Gaps   11;

Qy         33 KPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECAS 92
Db         61 KGHQLNSLPDQW----------YSGIVKPI-------LDELGLFNHLE--VYHTVDS 98

Qy         93 SNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYEL 152
Db         99 TQ---LKAKRALVGNKDTFLILSDEQTEGRGRFNRNWESSKGKGLWMSLVLRPDVPFSMI 155

Qy        153 GSLSPVAAVACRRALSRLGLD-VQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV-GIG 210
Db        156 PKFNLFIALGIRDAIQQFSNERVTIKWPNDIYIGNKKICGFLTEMVANYDEIEAIICGIG 215

Qy        211 INFVLPKEVENAASVQSLFQTASRRGNADAAV----LLETLLVELDAVLLQYARDGFAPF 266
Db        216 INM---NHVESDFDEDIKDKATSIRMHSDSIINRYTFLTALLTQIIHRFDQFLHQTFESI 272

Qy        267 VAEY-QAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK-QTVVSGEISL 322
Db        273 REEYIHATNIWHRQ--LKFTENNHQFLGEAIDIDSDGFLIVKDEKGQLHRLMSADIDL 328


RESULT 4
US-09-144-085-1
; Sequence 1, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 5087
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-144-085-1

  Query Match             3.8%;  Score 116;  DB 4;  Length 5087;
  Best Local Similarity   20.2%;  Pred. No. 0.042;
  Matches  137;  Conservative   86;  Mismatches  248;  Indels  206;  Gaps   33;

Qy         34 PQQLNGFWQQMPAHIRGLLRQHDGYWRL-------VRPLAV--FDAEGLRELGERSGFQT 84
Db       3240 PIAIVGVGLRMPGGANDL----ESFWQVLVEGRDTLRPIPTDRFDVEAMYDPNPEAKGKT 3295

Qy         85 ALKHECASSNDEILELAR----IAP------DKAHKTICVTHL----------QSKG-- 121
Db       3296 YVKH--ASLLDDVASFDAGFFGISPREAEPMDPQHRLLLETAWSALEDAGVRPDQLKGSD 3353

Qy        122 --------------RGRQGRKWSHRL-GECLMFSFGWVFDRPQYELGSLSPVAAV--AC 163
Db       3354 TGVFVGVAPSEYASYRGKSANEDAYALTGTALSFAAG----RVAYHLGLQGPAVSVDTAC 3409

Qy        164 RRALSRLGL--------DVQIKWPNDLVVGRDKLGGILIETVRTGG-----KTVAVV--- 207
Db       3410 SSSLVALHLACDALRRGDCEVALAAGVQVLANPAGFVLLSRTRAVSPDGRCKTFSQAADG 3469

Qy        208 ---GIGINFVLPKEVENAAS-----VQSLFQTASRRGNADAAVLLETLLVELDAVLLQYA 259
Db       3470 YGRGEGVGVVVLMRLSDAQAQGMRVLGVVRGTAVNQDGASSGITAPNGTAQQKVVRAALR 3529

Qy        260 RDGFAPFVAEYQAANRD----HGKAVLLLRDGETVFEGTVKGVDGQG------------ 302
Db       3530 NAGL-------EASSIDVVECHGTGTSL---GDPIEVQALGAVYGQGREATRPLRLGAVK 3579

Qy        303 --VLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
Db       3580 SNIGHLESAAGIAGVCKILAAFRHEALPATL---HSSPR-------NPQISWE------- 3622

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCG-----------EFKKAQVQEQLARK 409
```

```
Db       3623 ----SLPVQVVDRL-TGWPRRADGLPRFAGVSSFGISGTNAHVILEEAPLEAVREPAAVR 3677

Qy        410 IEWLPSSAQALGIRNHYRHPEEHG--SDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD 467

Db       3678 -EPLAAEGVAIPLLLSGRDEASVGAQAERWAKWLGEHAEVRWPDV--------------- 3721

Qy        468 DGHYLGGTIMPGFHLMKESLAVRTANLNR-HAGKRYPFPTTTGNAVASGMMDAVCGSVMM 526

Db       3722 --------------------VRTAALHRTHFAWR-------ASVQAASVSEAVEGLRAL 3753

Qy        527 MHGR-----LKEKTGAGKPVDVIITGGGAAKVAEA-----LPPAFLA--ENTVRVADNLV 574

Db       3754 SEGRAAAGVVRGTGGRGGKLAVLFTGQGSQRLGMGKRLYEVYPVFRAAFDEVCEALDAHL 3813

Qy        575 IYGLLNMIAAE-GREYE 590

Db       3814 DRGLREVVFAEAGSEQE 3830


RESULT 5
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Shah, Sanjay Krishnakant
;  APPLICANT: McDaniel, Robert
;  APPLICANT: Tang, Li
;  TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
;  FILE REFERENCE: 30062-20029.00
;  CURRENT APPLICATION NUMBER: US/09/428,517
;  CURRENT FILING DATE: 1999-10-28
;  EARLIER APPLICATION NUMBER: 60/120,254
;  EARLIER FILING DATE: 1999-02-16
;  EARLIER APPLICATION NUMBER: 60/106,100
;  EARLIER FILING DATE: 1998-10-29
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3

  Query Match             3.7%;  Score 114.5;  DB 4;  Length 3816;
  Best Local Similarity   20.4%;  Pred. No. 0.037;
  Matches  120;  Conservative   77;  Mismatches  213;  Indels  179;  Gaps   29;

Qy         96 EILELARIAPDKAHKTICVTHLQSKGR-------------GRQGRKWSHRLGECLMFSF 141

Db       1693 ELIESAGVAPDSLHRSRTGTFIGSNGQFYAPLLWNSGGDLEGYQG------VGNAGSVMS 1746

Qy        142 GWVFDRPQYELGSLSPVAAV--ACRRALSRLGLDVQ----------------IKWPNDL 182

Db       1747 G----RVAYSLGLEGPAVTVDTACSSSLVALHLAVQALRRGECSLAIAGGVTVMSTPDSF 1802

Qy        183 VV----------GRDK-------------LGGILIETV----RTGGKTVAVV-GIGINF 213

Db       1803 VEFSRQQGLSEDGRCKAFASTADGFGLAEGVSALLVERLSDARRLGHRVLAVVRGSAVN- 1861

Qy        214 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFA---PFVAEY 270

Db       1862 ------QDGASNGLTAPNGPSQQRVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQA 1915

Qy        271 QAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDDRPVS 330

Db       1916 LLATYGQGRA-----GGRPVVLGSVKSNIG----HTQAAAGVAGVMKMVLAL---ERGV- 1962

Qy        331 VPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVG 390

Db       1963 VPKT------LHVDEPSPHVDW-----------SAGEVELAVEAVPWSR--GGRVRRAG 2002

Qy        391 CA---VCGEFKKAQVQEQLARK--------------IEWLPSSAQALGIRNHYRHPEEH 432

Db       2003 VSSFGISGTNAHVIVEEAPAEPEPEPGTRVVAAGDLVVPWVVSGRDAGALREQAARLAAH 2062

Qy        433 GS-------DRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD-GHYLGGTIMPGFHLMK 484

Db       2063 VSSTGAGVVDVGWSLVATRSVFEHRAVMVGS----ELDSMAESLAGFAAGGVVPGV---- 2114

Qy        485 ESLAVRTANLNRHAGKR----YPFPTTTGNAVASGMMDAV---------CGSVMMMHGRL 531

Db       2115 ------VSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVL------ 2162

Qy        532 KEKTGAGKPVDVIITGGGAAKV--AEALPPAFLAENTVRVADNLVIYGL 578

Db       2163 --DPVTGWSLVEVLRGGGEAVLGRVDVVQPALWAV-MVSLARTWRYYGV 2208


RESULT 6
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 6095
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-144-085-2

  Query Match             3.6%;  Score 111.5;  DB 4;  Length 6095;
  Best Local Similarity   20.3%;  Pred. No. 0.19;
  Matches  155;  Conservative   84;  Mismatches  267;  Indels  259;  Gaps   37;

Qy          8 HWRVLAELADGLPQHVSQLARMA--DMKPQQLNGFWQQMPAHIR--GLLRQHDGYWRLVR 63

Db       5156 YWRLLAEGKDAIERFPSRYDAFSVYDPDPEAVG------KSYVREGGFLRDID------- 5202

Qy         64 PLAVFDAEGLRELGERSGFQTALKHECASSND----------EILELARIAPDKAHKTIC 113

Db       5203 ---VFDA----------GFFGISPREAQAMDPQQRLVLETAWEALERAGVRPSMLSESAT 5249

Qy        114 VTHLQSKGRGRQGRKWSHRLGECLMFSFGW---------VFDRPQYELGSLSPVAAV--A 162

Db       5250 GVYL-----GWMGSDYGALLGNDLAALDGYQGTGSAASVLSGRVAYVLGLQGPAITVDTA 5304

Qy        163 CRRALSRLGLDVQI--KWPNDLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVE 220

Db       5305 CSSSLVSLHLACTALRQGECDLAL----TGGVMVMTTPAG-----------FV---EFS 5345

Qy        221 NAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPF-VAEYQAANRD--- 276

Db       5346 RARGLARDGRCKSFSAQADGVIWSEGCGMLLLKRLSDARRDGDRVLGVIRGSAVNQDGRS 5405

Qy        277 --------------------------------HGKAVLLLRDGETVFEGTVKGVDG-- 300

Db       5406 QGLTAPNGPAQQRVIRQALSSCGLSPEDIDAVEAHGTGTSL---GDPIEAGALAEVFGPE 5462
```

```
Qy        301 ------------QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGN 347
                          : | : | |   |:   ::|: :  | ::   : |       |
Db       5463 RSPERPLYLGSSKSNLGHAQAAAGVAGVIKMVLALQHEVLPKTLHAEQPSPHIAWEGSGL 5522

Qy        348 SRLKWA--WVENG--------TFATVGSAPYRDLSPLGAEWAE---KADGNVRIVGCAVC 394
              | |: |  |  ||         :|   |:  :  |    ||       :|:    ::   :
Db       5523 SLLQEARPWRRNGRVRRAGVSSFGISGTNAHIILEEAPAEARREPVEAEAAPALLPLVLS 5582

Qy        395 GEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFN-----ALGSRRF-SR 448
              |   :| |  |  |  :||                  ||||   | :     ||    | ||
Db       5583 GR-DEASVAAQAGRWAKWL---------------EEHGEVGWSDVVRTAALHRTHFESR 5625

Qy        449 NACVVVSCGTAVTV-DALTDD-GH----------------YLG-GTIMPGF--HLMKES 486
               : :  |    | |  ||::  ||                : | |:  ||    |:::|
Db       5626 ASMLAASVSEVVEVLRALSEGRGHRAVSVGTARARGKVVFVFPGQGSQWPGMGRALLEQS 5685

Qy        487 LAVRTA----------------NLNRHAGKRYP-----------FPTTTGNAVA--- 513
               |   |                |    |:  |           |    | | |
Db       5686 AAFAEAVQACDEALRPWTGWSVLSVLRGDGGEEQPSLERVDVVQPALFAMCVGLAAAWRS 5745

Qy        514 -----------SGMMDA--VCGSVMMMHG---------RLKEKTGAGKPVDVIITGGGA 550
                         | : |  |||:: :  |          :::::| |    :::
Db       5746 LGLEPAAVVGHSQGEVSAAVVCGALSLAEGARVVALRSQAVRQQSGMGA---MMLVEQPV 5802

Qy        551 AKVAEALPP-----AFLAEN----TVRVADNLVIYGLLNMIAAEG 586
              ::| | : |     |  | |    ||   |   : ||:  : |||
Db       5803 SEVQERIAPYGEALAIAAVNTSNSTVVSGDVEAVDGLMVELTAEG 5847
```

RESULT 7
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow AgroSciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5588 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-6

Query Match 3.5%; Score 109; DB 4; Length 5588;
Best Local Similarity 18.5%; Pred. No. 0.31;
Matches 150; Conservative 81; Mismatches 268; Indels 314; Gaps 32;

```
Qy         10 RVLAELADGLPQHVSQLARMAD------MKPQQLNGFWQQMPAHIRGLLRQHDGYW---- 59
              ||  : |   ||       |: |       :| ::  |:  :        |    ||
Db        747 RVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAEI-AFYSTVTGERTDTGRLDADYWYQNL 805

Qy         60 ----RLVRPLAVFDAEGLRELGERS-------GFQTALKHECASSNDEILELARIAPDKA 108
                  | : :|    :| |   | |       | |  |  | | :   ::   |
Db        806 RQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGIQETL--EAADAGGVVVGSLR------ 857

Qy        109 HKTICVTHLQSKGRGRQG-RKWSHRLGEC----LMFSFGWVF------------------ 145
                        ||  | |:|   | ||    |  ::  ||
Db        858 -------------RGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGARRVPLPTYPFQRQ 904

Qy        146 ----DRPQYELGSLSPVAAVACRRALSRLGLDVQI----------------KWPNDLVV 184
                  :  :|: | |  |  ::    |  ||  | :                 |  | ||
Db        905 RYWLESAEYDAGDLGSVGLLSAEHPL--LGAAVTLADAGGFLLTGKLSVKTQPWLADHVV 962

Qy        185 G----------------RDKLGGILIETV----------------------RTGGKT 203
              |                  |::|  ||| :                      | ::
Db        963 GGAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRS 1022

Qy        204 VAV---------------------------VGIGINFVLPKEVENAASVQSLFQTASRR 235
              | |                           | | : : |     |  : | :  |: |
Db       1023 VRVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPLDSFYARAAER 1082

Qy        236 G-----------------------------NADAA------VLLETLLVELDAVLLQYA 259
              |                             ||       ||:  |  | |   :
Db       1083 GFDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQALGAA-EEDP 1141

Qy        260 RDGFAPFVAEYQAANRDHGKA----VLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTV 315
               :|: ||   :|  :     |    | |:  |             || || ::      ::
Db       1142 DEGWLPFA--WQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSID------SL 1193

Qy        316 VSGEISLRSDDRPVSVPKRRDSERFLLLD------GGNSRLKWAWVENGTFA-TVGSAPY 368
              |  :||    |: ::  :  : |   :|      |    : || : |   |   |||
Db       1194 VLRQIS----DKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALG 1249

Qy        369 RDLSPLGAEWAEKAD--GNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHY 426
               :| |   :    ||   :| :| |                          :| | | : :
Db       1250 TELHP---DLTGLADPPPDVVVVPCG---------------------ASRQDLDVASEA 1284

Qy        427 RHPEEHGSDRWFNALGSRRFSRNACVVVSCGTA-------------------------- 459
              |   :   |  : | : ||: :  |||:|| |
Db       1285 RAATQRMLDLIQDWLAAARFAGSRLVVVTCGAASTGPAEGVSDLVHAASWGLLRSAQSEN 1344

Qy        460 ------VTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN--------RHAGKRYPFP 505
                  | ||   :    |  :  |    :  ||:|   :                | | |
Db       1345 PDRFVLVDVDGTAESWRALAAAVRSG----EPQLALRAGEVRVPRLARCVAAEDSRIPVP 1400

Qy        506 TTTGNAVASGMMDAVCGSV-----------MMMHGRLK-EKTGAGKPVDVIITGGGAAK 552
                 |  : ||    : | |            ::: ||       |    || ::  | | :
Db       1401 GADGTVLISGGTGLLGGLVARHLVAERGVRRLVLAGRRGWSAPGVTDLVDELVGLGAAVE 1460

Qy        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAE 585
              ||          :  | |   :  ||  |:||
Db       1461 VA----------SCDVGDRAQLDRLLTTISAE 1482
```

RESULT 8
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy

```
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match             3.5%;  Score 109;  DB 4;  Length 5588;
  Best Local Similarity   18.5%;  Pred. No. 0.31;
  Matches  150;  Conservative   81;  Mismatches  268;  Indels  314;  Gaps   32;

Qy         10 RVLAELADGLPQHVSQLARMAD------MKPQQLNGFWQQMPAHIRGLLRQHDGYW---- 59
              ||  : |   ||      |: |       :| ::  |:  :         |    ||
Db        747 RVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAEI-AFYSTVTGERTDTGRLDADYWYQNL 805

Qy         60 ----RLVRPLAVFDAEGLRELGERS-------GFQTALKHECASSNDEILELARIAPDKA 108
                  | : :|    :| |   | |       | |  |  | | :   ::   |
Db        806 RQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGIQETL--EAADAGGVVVGSLR------ 857

Qy        109 HKTICVTHLQSKGRGRQG-RKWSHRLGEC----LMFSFGWVF----------------- 145
                     ||  | |:|   | ||    |  ::  ||
Db        858 ------------RGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGARRVPLPTYPFQRQ 904

Qy        146 ----DRPQYELGSLSPVAAVACRRALSRLGLDVQI----------------KWPNDLVV 184
                  :  :|: | |  |  ::    |  ||  | :                 |  | ||
Db        905 RYWLESAEYDAGDLGSVGLLSAEHPL--LGAAVTLADAGGFLLTGKLSVKTQPWLADHVV 962

Qy        185 G----------------RDKLGGILIETV----------------------RTGGKT 203
              |                |:::|  ||| :                      | ::
Db        963 GGAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRS 1022

Qy        204 VAV--------------------------VGIGINFVLPKEVENAASVQSLFQTASRR 235
              | |                          | | : : |    |  : | :  |: |
Db       1023 VRVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPLDSFYARAAER 1082

Qy        236 G----------------------------NADAA------VLLETLLVELDAVLLQYA 259
              |                            ||       ||:  |  | |   :
Db       1083 GFDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQALGAA-EEDP 1141

Qy        260 RDGFAPFVAEYQAANRDHGKA----VLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTV 315
               :|: ||   :|  :      |    | |:  |             || || ::      ::
Db       1142 DEGWLPFA--WQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSID------SL 1193

Qy        316 VSGEISLRSDDRPVSVPKRRDSERFLLLD------GGNSRLKWAWVENGTFA-TVGSAPY 368
              |  :||    |: ::  :  : |    :|      |    : || : |   |   |||
Db       1194 VLRQIS----DKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALG 1249

Qy        369 RDLSPLGAEWAEKAD--GNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHY 426
               :| |   :    ||   :| :| |                        :| | | : :
Db       1250 TELHP---DLTGLADPPPDVVVVPCG--------------------ASRQDLDVASEA 1284

Qy        427 RHPEEHGSDRWFNALGSRRFSRNACVVVSCGTA------------------------- 459
              |   :   |   : | : ||: :  |||:|| |
Db       1285 RAATQRMLDLIQDWLAAARFAGSRLVVVTCGAASTGPAEGVSDLVHAASWGLLRSAQSEN 1344

Qy        460 ------VTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN--------RHAGKRYPFP 505
                   | ||   :    |  : |   :  ||:|   :                | | |
Db       1345 PDRFVLVDVDGTAESWRALAAAVRSG----EPQLALRAGEVRVPRLARCVAAEDSRIPVP 1400

Qy        506 TTTGNAVASGMMDAVCGSV----------MMMHGRLK-EKTGAGKPVDVIITGGGAAK 552
               |  : ||    : | |            ::: ||      |    || ::  | | :
Db       1401 GADGTVLISGGTGLLGGLVARHLVAERGVRRLVLAGRRGWSAPGVTDLVDELVGLGAAVE 1460

Qy        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAE 585
                ||           :  | |   :  ||  |:||
Db       1461 VA----------SCDVGDRAQLDRLLTTISAE 1482


RESULT 9
US-07-642-734C-5
; Sequence 5, Application US/07642734C
; Patent No. 5824513
;  GENERAL INFORMATION:
;    APPLICANT:  Katz, L
;    APPLICANT:  Donadio, S
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Recombinant DNA Method for Producing
;    TITLE OF INVENTION:  Erythromycin Analogs
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Edward H. Gorman
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/642,734C
;      FILING DATE:  17-JAN-91
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Danckers, Andreas M
;      REGISTRATION NUMBER:  32652
;      REFERENCE/DOCKET NUMBER:  4952.US.01
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  708-937-9396
;      TELEFAX:  708-938-2623
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3170 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-642-734C-5

  Query Match             3.5%;  Score 108.5;  DB 2;  Length 3170;
  Best Local Similarity   22.5%;  Pred. No. 0.13;
  Matches  129;  Conservative   67;  Mismatches  232;  Indels  145;  Gaps   30;

Qy         46 AHIRGLLR----QHDGYWRLVRPLAVFDAEGLRELG-----ERSGFQTALKHECASSNDE 96
              | : || |     :  | |  :  |    ||  :| |       |   | |:: :
Db       1041 AMVGGLGRVVGIEQPGRWGGLVDLVDADAASIRSLAAVLADPRGEEQVAIRADG------ 1094

Qy         97 ILELARIAPDKAHK---------TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDR 147
                 :::||: |  |           |: ||   : | |     :|  | |    :  |
Db       1095 -IKVARLVPAPARAARTRWSPRGTVLVTG-GTGGIGAHVARWLARSGAEHLVLLG----- 1147

Qy        148 PQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV 207
                   |: :|  |   |  |: ||  | |   :  |  | :|  :|    |  |:||: |
Db       1148 ---RRGADAP-GASELREELTALGTGVTIAACD--VADRARLEAVL-AAERAEGRTVSAV 1200

Qy        208 --GIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGF-- 263
                  |::   | :   |   :     |      | |: |  :||| :|  :  |
Db       1201 MHAAGVSTSTPLDDLTEAEFTEIADVKVR-----GTVNLDELCPDLDAFVLFSSNAGVWG 1255

Qy        264 APFVAEYQAAN----------RDHGKAVLLLRDGETVFEG-TVKGVDGQGVLHLETAEGK 312
              :| :| | |||          |  |  |  :  |  :: |  : | :|
Db       1256 SPGLASYAAANAFLDGFARRRRSEGAPVTSIAWG--LWAGQNMAGDEG------------ 1301

Qy        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAP----- 367
```

```
Db       1302 -----GEY-LRSQGLRAMDPDRAVEELHITLDHGQTSVSVVDMDRRRFVELFTAARHRPL 1355

Qy        368 YRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYR 427

Db       1356 FDEIAGARAEARQSEEGPALAQRLAA---LSTAERREHLAHLIR--AEVAAVLG------ 1404

Qy        428 HPEEHGS------DRWFNALGSRRFS----RNACVVVS----CGTAV----TVDALTDDG 469

Db       1405 ----HGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAAVTGVREAATVVFDHPTITRLAD-- 1458

Qy        470 HYLGGTI-------MPGFHLMKE------------SLAVRTANLNRHAGKRYPFPTTTGN 510

Db       1459 HYLERLVGAAEAEQAPA--LVREVPKDADDPIAIVGMACRFPGGVHNPGELWEFIVGRGD 1516

Qy        511 AVA------SGMMDAVCGSVMMMHGRLKEKTGA 537

Db       1517 AVTEMPTDRGWDLDALFDPDPQRHGTSYSRHGA 1549


RESULT 10
 S-08-439-009A-5
; Sequence 5, Application US/08439009A
; Patent No. 6004787
;  GENERAL INFORMATION:
;    APPLICANT:  Donadio, S
;    APPLICANT:  Katz, L
;    APPLICANT:  Mcalpine, J B
;    TITLE OF INVENTION:  Method of Directing Biosynthesis of
;    TITLE OF INVENTION:  Specific Polyketides
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Steven F. Weinstock
;      STREET:  Abbott Laboratories D377/AP6D-2 One Abbott
;      STREET:  Park Rd
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/439,009A
;      FILING DATE:  11-MAY-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Casuto, Dianne
;      REGISTRATION NUMBER:  40,943
;      REFERENCE/DOCKET NUMBER:  4952.US.D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847-938-3137
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3170 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-439-009A-5

  Query Match             3.5%;  Score 108.5;  DB 3;  Length 3170;
  Best Local Similarity   22.5%;  Pred. No. 0.13;
  Matches  129;  Conservative   67;  Mismatches  232;  Indels  145;  Gaps   30;

Qy         46 AHIRGLLR----QHDGYWRLVRPLAVFDAEGLRELG-----ERSGFQTALKHECASSNDE 96

Db       1041 AMVGGLGRVVGIEQPGRWGGLVDLVDADAASIRSLAAVLADPRGEEQVAIRADG------ 1094

Qy         97 ILELARIAPDKAHK---------TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDR 147

Db       1095 -IKVARLVPAPARAARTRWSPRGTVLVTG-GTGGIGAHVARWLARSGAEHLVLLG----- 1147
```

```
Qy        148 PQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV 207

Db       1148 ---RRGADAP-GASELREELTALGTGVTIAACD--VADRARLEAVL-AAERAEGRTVSAV 1200

Qy        208 --GIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGF-- 263

Db       1201 MHAAGVSTSTPLDDLTEAEFTEIADVKVR-----GTVNLDELCPDLDAFVLFSSNAGVWG 1255

Qy        264 APFVAEYQAAN----------RDHGKAVLLLRDGETVFEG-TVKGVDGQGVLHLETAEGK 312

Db       1256 SPGLASYAAANAFLDGFARRRRSEGAPVTSIAWG--LWAGQNMAGDEG------------ 1301

Qy        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAP----- 367

Db       1302 -----GEY-LRSQGLRAMDPDRAVEELHITLDHGQTSVSVVDMDRRRFVELFTAARHRPL 1355

Qy        368 YRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYR 427

Db       1356 FDEIAGARAEARQSEEGPALAQRLAA---LSTAERREHLAHLIR--AEVAAVLG------ 1404

Qy        428 HPEEHGS------DRWFNALGSRRFS----RNACVVVS----CGTAV----TVDALTDDG 469

Db       1405 ----HGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAAVTGVREAATVVFDHPTITRLAD-- 1458

Qy        470 HYLGGTI-------MPGFHLMKE------------SLAVRTANLNRHAGKRYPFPTTTGN 510

Db       1459 HYLERLVGAAEAEQAPA--LVREVPKDADDPIAIVGMACRFPGGVHNPGELWEFIVGRGD 1516

Qy        511 AVA-----SGMMDAVCGSVMMMHGRLKEKTGA 537

Db       1517 AVTEMPTDRGWDLDALFDPDPQRHGTSYSRHGA 1549


RESULT 11
5310667-5
;Patent No. 5310667
;    APPLICANT: EICHHOLTZ, DAVID A.;GASSER, CHARLES S.;KISHORE,
;GANESH M.
;    TITLE OF INVENTION: GLYPHOSATE-TOLERANT 5-ENOLPYRUVYL
;-3-PHOSPHOSHIKIMATE SYNTHASES
;    NUMBER OF SEQUENCES: 37
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/380,963
;      FILING DATE: 17-JUL-1989
;SEQ ID NO:5:
;      LENGTH: 444
5310667-5

  Query Match             3.5%;  Score 107;  DB 6;  Length 444;
  Best Local Similarity   24.8%;  Pred. No. 0.0059;
  Matches   76;  Conservative   34;  Mismatches  101;  Indels   96;  Gaps   18;

Qy        166 ALSRLGLDVQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASV 225

Db         58 ALRTLGLSVE----------ADK----------AAKRAVVVGCGGKF----PVEDAKEE 92

Qy        226 QSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLR 285

Db         93 VQLF-----LGNAGTAMRPLTAAVTAAGGNATYVLDG-VPRMRE-----RPIGDLVVGLK 141

Qy        286 D-GETV--FEGT------VKGVDG--QGVLHLETAEGKQ---------TVVSGEISLRSD 325

Db        142 QLGADVDCFLGTDCPPVRVNGIGGLPGGKVKLSGSISSQYLSALLMAAPLALGDVEIEII 201

Qy        326 DRPVSVP----------------KRRDS-ERFLLLDGGNSRL-KWAWVE----NGTFATV 363

Db        202 DKLISIPYVEMTLRLMERFGVKAEHSDSWDRFYIKGGQKYKSPKNAYVEGDASSATYFLA 261

Qy        364 GSAPYRDLSPLGAEWAEKADGNVRIVGC---AVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Db        262 GAA--------------ITGGTVTVEGCGTTSLQGDVKFAEVLEMMGAKVTWTETSVTVT 307
```

```
Qy        421 G-IRNHY 426
              |  |:|:
Db        308 GPPRSHF 314


RESULT 12
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             3.4%;  Score 105.5;  DB 4;  Length 4613;
  Best Local Similarity   22.4%;  Pred. No. 0.54;
  Matches  104;  Conservative   52;  Mismatches  174;  Indels  135;  Gaps   22;

Qy        142 GWVFDRPQYELGSLSPVAAV------------AC---RRALSRLGL--DVQIKWPNDLV 183
              | : :|  | ||   |   |                ||   ||  | | |   | :   | :
Db        237 GIIANRLSYTLGLRGPSMVVDSGQSSSLVAVHLACESLRRGESELALAGGVSLNLVPDSI 296

Qy        184 VGRDKLGGILIE------TVRTGGKTVAVVGIGINFVLPKEVENAAS----VQSLFQTAS 233
              :|  | ||:  :        |  |    | | |  ||: | :  | :    | :: :  |
Db        297 IGASKFGGLSPDGRAYTFDARANG---YVRGEGGGFVVLKRLSRAVADGDPVLAVIR-GS 352

Qy        234 RRGNADAAVLLETLLVELDAVLLQYA--RDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                | ||  : |   :    :|: |  | | ||    |      ||    :   |
Db        353 AVNNGGAAQGMTTPDAQAQEAVLREAHERAGTAPADVRYVEL---HGTGTPVGDPIEAAA 409

Qy        292 EGTVKGVD---GQGVL---------HLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSER 339
              |   |     || :|         ||| | |   ::   :::|    | |:
Db        410 LGAALGTGRPAGQPLLVGSVKTNIGHLEGAAGIAGLIKAVLAVRGRALPASL-------- 461

Qy        340 FLLLDGGNSRLKWAWVENGTFATVGSA-PYRDLS-PLGAE---WAEKADGNVRIVGCAVC 394
                          : |   | |: :|:  :  |   |  : ||   :|| :
Db        462 ------------------NYETPNPAIPFEELNLRVNTEYLPWEPEHDGQRMVVGVSSF 502

Qy        395 GEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVV 454
              |                        :|  | :   ||        | |     | | ||
Db        503 G------------------------MGGTNAHVVLEE--------APGG---CRGASVVE 527

Qy        455 SCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHA--GKRYPFPTTTGNAV 512
              |          |  |  :|| ::| : :  :| |   | : | |    |         ||
Db        528 S----------TVGGSAVGGGVVP-WVVSAKSAAALDAQIERLAAFASRDRTDGVDAGAV 576

Qy        513 ASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
              :| :||   : ::  || : :  |      :: | |   :| ||
Db        577 DAGAVDAGAVARVLAGGRAQFEHRA------VVVGSGPDDLAAAL 615


RESULT 13
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             3.4%;  Score 105.5;  DB 4;  Length 11877;
  Best Local Similarity   22.4%;  Pred. No. 2.9;
  Matches  104;  Conservative   52;  Mismatches  174;  Indels  135;  Gaps   22;

Qy        142 GWVFDRPQYELGSLSPVAAV------------AC---RRALSRLGL--DVQIKWPNDLV 183
              | : :|  | ||   |   |                ||   ||  | | |   | :   | :
Db        573 GIIANRLSYTLGLRGPSMVVDSGQSSSLVAVHLACESLRRGESELALAGGVSLNLVPDSI 632

Qy        184 VGRDKLGGILIE------TVRTGGKTVAVVGIGINFVLPKEVENAAS----VQSLFQTAS 233
              :|  | ||:  :        |  |    | | |  ||: | :  | :    | :: :  |
Db        633 IGASKFGGLSPDGRAYTFDARANG---YVRGEGGGFVVLKRLSRAVADGDPVLAVIR-GS 688

Qy        234 RRGNADAAVLLETLLVELDAVLLQYA--RDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                | ||  : |   :    :|: |  | | ||    |      ||    :   |
Db        689 AVNNGGAAQGMTTPDAQAQEAVLREAHERAGTAPADVRYVEL---HGTGTPVGDPIEAAA 745

Qy        292 EGTVKGVD---GQGVL---------HLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSER 339
              |   |     || :|         ||| | |   ::   :::|    | |:
Db        746 LGAALGTGRPAGQPLLVGSVKTNIGHLEGAAGIAGLIKAVLAVRGRALPASL-------- 797

Qy        340 FLLLDGGNSRLKWAWVENGTFATVGSA-PYRDLS-PLGAE---WAEKADGNVRIVGCAVC 394
                          : |   | |: :|:  :  |   |  : ||   :|| :
Db        798 ------------------NYETPNPAIPFEELNLRVNTEYLPWEPEHDGQRMVVGVSSF 838

Qy        395 GEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVV 454
              |                        :|  | :   ||        | |     | | ||
Db        839 G------------------------MGGTNAHVVLEE--------APGG---CRGASVVE 863

Qy        455 SCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHA--GKRYPFPTTTGNAV 512
              |          |  |  :|| ::| : :  :| |   | : | |    |         ||
Db        864 S----------TVGGSAVGGGVVP-WVVSAKSAAALDAQIERLAAFASRDRTDGVDAGAV 912

Qy        513 ASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
              :| :||   : ::  || : :  |      :: | |   :| ||
Db        913 DAGAVDAGAVARVLAGGRAQFEHRA------VVVGSGPDDLAAAL 951


RESULT 14
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4928 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
S-09-036-987A-5

  Query Match             3.4%;  Score 105;  DB 4;  Length 4928;
  Best Local Similarity   21.5%;  Pred. No. 0.69;
  Matches  176;  Conservative   74;  Mismatches  261;  Indels  308;  Gaps   45;

Qy         11 VLAELADGLPQ--------HVSQLARMADMKPQQLNGFWQ----------QMPAHIRGLL 52
              :| |:|| |           ||       |:|  |  | |          ::|  |  |:
Db        745 LLTEIADRLHDQGCRTRWLRVSHAFHSPHMEP-MLEEFAQISRGREYHAPELPI-ISTLI 802

Qy         53 RQHDG--------YW-RLVRPLAVFDAEGLR----------ELGERSGFQTALKHECAS 92
               : ||        || | ||    | |||::           |||   |  : |  || :
Db        803 GELDGGRVMGTPEYWVRQVREPVRF-AEGVQALVGQGVGTIVELGP-DGALSTLVEECVA 860

Qy         93 SNDEI--LELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSF--------- 141
              :  :  : | |   |:| :|:     |   ||          ||    ||
Db        861 ESGRVAGIPLMRKDRDEA-RTVLAALAQIHTRG----------GEVDWRSFFAGTGAKQV 909

Qy        142 ---GWVFDRPQYELGS-----------------LSPVAAVACRRALSRLG--LDVQIK 177
                 : | | :| | |                  |  | |:|    :   |
Db        910 DLPTYAFQRQRYWLASTGRAGDVTAAGLAEADHPLLGAVVALADGEGVVLTGRLTAGSHP 969

Qy        178 WPND-LVVGRDKL-GGILIETV-----RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQ 230
              | :|  |:|   : |  ::| |     | |   |  : :   :||     |  || |
Db        970 WLSDHRVLGEIVVPGTAIVELVWHVGERLGCGRVEELALEAPLILPD--HGAVQVQVLVG 1027

Qy        231 TASRRGNADAA----------------VLLETL-------------LVELDAVLLQ 257
                |   |                   |||  :              |:||
Db       1028 PPGESGARSVALYSCPGEAIEPEWKKHATGVLLPPVAAENHELTAWPPENATEIDA---- 1083

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRD----GETVF------EGTVKGVDGQG----- 302
                ||   |:  :  |   :| |   ||      |  ||       :    |||  |
Db       1084 ---DGVYAFLEGHGFA---YGPAFRCLRGAWRRGGEVFAEVALPDDMQAGVDRFGVHPAL 1137

Qy        303 ---VLHLETAE--------------------GKQTVVSGEISLRSDD------------ 326
                 |||   ||                      :  ||   :|| |||
Db       1138 LDAVLHAAAAETSVVQSEARVPFSWRGVELRATESAVVRARLSLTSDDELSLVAVDPAGR 1197

Qy        327 ----------RPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPL-- 374
                     ||:|  : |          |:   :  |       | :|:    ||: :
Db       1198 FVATVDSLVTRPISRQQVRSGAI------GDCLFEVEWHRK---ALLGTTAGDDLAIVGD 1248

Qy        375 GAEWAEKADGNVRIVGCAVCGEFKKA----------QVQEQLARKIEW--LPSSAQALG 421
              |  | |      |    |   ||: |          |    | ::|   ||| ||
Db       1249 GPSWPESVRATARF---ATLDEFRAAVDSDVPAPGSVLVAAMSAEEVEGGSLPSRAQ--- 1302

Qy        422 IRNHYRHPEEHGSD------RWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 475
                  |  ||       |   |   ||: :  |||:   ||: |: :|    :| :
Db       1303 ---------ESTSDLLALVQSW---LADERFAESQLVVVT-RAAVSADSDSDVADLVGAS 1349

Qy        476 IM----------PGFHLMKE--------------------------SLAVRTANLNRH 497
                         ||  :: :                         :|  | | |
Db       1350 SWGLLSSAQSENPGRFVLVDVDGTPESWQALPAAVRAGEPQLALRRGVALVPRLARLTVR 1409

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
                  |  | |  : :|   |: | |:  |  | |: |  : |     |  | | |
Db       1410 EEGSSPQLDTDGTVLITGGTGAL-GGVVARH--LVEEHGIRRLVLAGRRGWNAPGVHE-- 1464

Qy        558 PPAFLAENTVRVADNLVIYG-LLNMIA---AEGREYEHI 592
                        : | |   | :: ::|   |:  : ||:
Db       1465 ----------LVDELARAGAVVEVVACDVADRTDLEHV 1492


RESULT 15
US-09-370-700-5
; Sequence 5, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 4928
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-5

  Query Match             3.4%;  Score 105;  DB 4;  Length 4928;
  Best Local Similarity   21.5%;  Pred. No. 0.69;
  Matches  176;  Conservative   74;  Mismatches  261;  Indels  308;  Gaps   45;

Qy         11 VLAELADGLPQ--------HVSQLARMADMKPQQLNGFWQ----------QMPAHIRGLL 52
              :| |:|| |           ||       |:|  |  | |          ::|  |  |:
Db        745 LLTEIADRLHDQGCRTRWLRVSHAFHSPHMEP-MLEEFAQISRGREYHAPELPI-ISTLI 802

Qy         53 RQHDG--------YW-RLVRPLAVFDAEGLR----------ELGERSGFQTALKHECAS 92
               : ||        || | ||    | |||::           |||   |  : |  || :
Db        803 GELDGGRVMGTPEYWVRQVREPVRF-AEGVQALVGQGVGTIVELGP-DGALSTLVEECVA 860

Qy         93 SNDEI--LELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSF--------- 141
              :  :  : | |   |:| :|:     |   ||          ||    ||
Db        861 ESGRVAGIPLMRKDRDEA-RTVLAALAQIHTRG----------GEVDWRSFFAGTGAKQV 909

Qy        142 ---GWVFDRPQYELGS-----------------LSPVAAVACRRALSRLG--LDVQIK 177
                 : | | :| | |                  |  | |:|    :   |
Db        910 DLPTYAFQRQRYWLASTGRAGDVTAAGLAEADHPLLGAVVALADGEGVVLTGRLTAGSHP 969

Qy        178 WPND-LVVGRDKL-GGILIETV-----RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQ 230
              | :|  |:|   : |  ::| |     | |   |  : :   :||     |  || |
Db        970 WLSDHRVLGEIVVPGTAIVELVWHVGERLGCGRVEELALEAPLILPD--HGAVQVQVLVG 1027

Qy        231 TASRRGNADAA----------------VLLETL-------------LVELDAVLLQ 257
                |   |                   |||  :              |:||
Db       1028 PPGESGARSVALYSCPGEAIEPEWKKHATGVLLPPVAAENHELTAWPPENATEIDA---- 1083

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRD----GETVF------EGTVKGVDGQG----- 302
                ||   |:  :  |   :| |   ||      |  ||       :    |||  |
Db       1084 ---DGVYAFLEGHGFA---YGPAFRCLRGAWRRGGEVFAEVALPDDMQAGVDRFGVHPAL 1137

Qy        303 ---VLHLETAE--------------------GKQTVVSGEISLRSDD------------ 326
                 |||   ||                      :  ||   :|| |||
Db       1138 LDAVLHAAAAETSVVQSEARVPFSWRGVELRATESAVVRARLSLTSDDELSLVAVDPAGR 1197

Qy        327 ----------RPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPL-- 374
```

```
                    ||:|  : |          |:   :  |      | :|:    ||: :
Db       1198 FVATVDSLVTRPISRQQVRSGAI------GDCLFEVEWHRK---ALLGTTAGDDLAIVGD 1248

Qy        375 GAEWAEKADGNVRIVGCAVCGEFKKA-----------QVQEQLARKIEW--LPSSAQALG 421
             |  | |       |    |   ||: |            |    | ::|   ||| ||
Db       1249 GPSWPESVRATARF---ATLDEFRAAVDSDVPAPGSVLVAAMSAEEVEGGSLPSRAQ--- 1302

Qy        422 IRNHYRHPEEHGSD------RWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 475
                 |  ||       |   |   ||: :  |||:   ||: |: :|    :| :
Db       1303 ---------ESTSDLLALVQSW---LADERFAESQLVVVT-RAAVSADSDSDVADLVGAS 1349

Qy        476 IM----------PGFHLMKE--------------------------SLAVRTANLNRH 497
                  ||  :: :                                :|  | | |
Db       1350 SWGLLSSAQSENPGRFVLVDVDGTPESWQALPAAVRAGEPQLALRRGVALVPRLARLTVR 1409

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
               |   | |  : :|   |: | |:  |  | |: |  : |     |  |  | |
Db       1410 EEGSSPQLDTDGTVLITGGTGAL-GGVVARH--LVEEHGIRRLVLAGRRGWNAPGVHE-- 1464

Qy        558 PPAFLAENTVRVADNLVIYG-LLNMIA---AEGREYEHI 592
                 : | |   | :: ::|   |:  : ||:
Db       1465 ----------LVDELARAGAVVEVVACDVADRTDLEHV 1492
```

Search completed: June 24, 2003, 21:49:45
Job time : 21.2942 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 48.2837 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-43
Perfect score: 3078
Sequence: 1 MTVLKLSHWRVLAELADGLP.....LVIYGLLNMIAAEGREYEHI 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3078	100.0	592	16	Q9JXF1	Q9jxf1 neisseria m
2	3024	98.2	592	16	Q9JWI7	Q9jwi7 neisseria m
3	324	10.5	291	16	Q8Y2M3	Q8y2m3 ralstonia s
4	308.5	10.0	321	16	Q8X709	Q8x709 escherichia
5	294.5	9.6	320	16	Q9KV39	Q9kv39 vibrio chol
6	293	9.5	288	2	Q9ZIU6	Q9ziu6 bordetella
7	284.5	9.2	295	16	Q8Y2M4	Q8y2m4 ralstonia s
8	279.5	9.1	312	16	Q9HWC0	Q9hwc0 pseudomonas
9	275.5	9.0	320	16	Q8Z317	Q8z317 salmonella
10	274.5	8.9	319	16	Q8ZAN5	Q8zan5 yersinia pe
11	244	7.9	322	2	O32511	O32511 deinococcus
12	244	7.9	474	16	Q9RRZ1	Q9rrz1 deinococcus
13	238	7.7	312	16	Q9CNX6	Q9cnx6 pasteurella
14	235.5	7.7	317	16	Q9PCI3	Q9pci3 xylella fas
15	234.5	7.6	248	16	Q9HWC1	Q9hwc1 pseudomonas
16	234	7.6	326	16	Q8R8N0	Q8r8n0 thermoanaer
17	228	7.4	242	2	Q52660	Q52660 rhodobacter
18	223.5	7.3	274	16	Q9CCL3	Q9ccl3 mycobacteri
19	221	7.2	255	16	Q92QN8	Q92qn8 rhizobium m
20	221	7.2	257	2	Q9R9L1	Q9r9l1 rhizobium m
21	218.5	7.1	325	16	Q92AA4	Q92aa4 listeria in
22	218	7.1	323	16	Q9KC88	Q9kc88 bacillus ha
23	217.5	7.1	325	16	Q8Y5Z9	Q8y5z9 listeria mo
24	208.5	6.8	274	16	P72816	P72816 synechocyst
25	208	6.8	266	16	P96884	P96884 mycobacteri
26	203	6.6	233	16	O66837	O66837 aquifex aeo
27	203	6.6	235	17	O73952	O73952 pyrococcus
28	202.5	6.6	261	17	O27938	O27938 methanobact
29	200	6.5	242	16	Q9PCI4	Q9pci4 xylella fas
30	198	6.4	166	5	Q9N802	Q9n802 leishmania
31	198	6.4	333	1	Q977S0	Q977s0 uncultured
32	195.5	6.4	237	17	Q9V2C7	Q9v2c7 pyrococcus
33	195	6.3	235	17	O57883	O57883 pyrococcus
34	194	6.3	311	16	Q97NV9	Q97nv9 streptococc
35	193	6.3	291	16	Q9EWV9	Q9ewv9 streptomyce
36	190.5	6.2	323	16	Q99U33	Q99u33 staphylococ
37	188.5	6.1	325	17	Q8TSW9	Q8tsw9 methanosarc
38	187.5	6.1	326	1	Q9HH16	Q9hh16 methanosarc
39	184.5	6.0	229	1	Q977N1	Q977n1 uncultured
40	183.5	6.0	250	16	Q9A6Z0	Q9a6z0 caulobacter
41	181	5.9	273	16	Q97EB4	Q97eb4 clostridium
42	179.5	5.8	231	16	Q9WY67	Q9wy67 thermotoga
43	177	5.8	265	16	Q9X8N6	Q9x8n6 streptomyce
44	175.5	5.7	256	16	Q8UFW3	Q8ufw3 agrobacteri
45	174.5	5.7	268	16	Q9ZD17	Q9zd17 rickettsia

ALIGNMENTS

RESULT 1
Q9JXF1
ID Q9JXF1 PRELIMINARY; PRT; 592 AA.
AC Q9JXF1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BirA protein/Bvg accessory factor.
GN NMB2075.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002557; AAF42394.1; -.
DR HSSP; P06709; 1BIA.
DR TIGR; NMB2075; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BirA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00121; birA_ligase; 1.

```
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  64701 MW;  9E27421DA2B41DE3 CRC64;

  Query Match           100.0%;  Score 3078;  DB 16;  Length 592;
  Best Local Similarity  100.0%;  Pred. No. 4.1e-209;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592


RESULT 2
Q9JWI7
ID   Q9JWI7      PRELIMINARY;      PRT;   592 AA.
AC   Q9JWI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein (EC 6.3.4.15).
GN   BIRA OR NMA0357.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162752; CAB83659.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   592 AA;  64468 MW;  7ED2E0CD3B31C630 CRC64;

  Query Match           98.2%;  Score 3024;  DB 16;  Length 592;
  Best Local Similarity  98.5%;  Pred. No. 2.7e-205;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 3
Q8Y2M3
ID   Q8Y2M3      PRELIMINARY;      PRT;   291 AA.
AC   Q8Y2M3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative bifunctional protein: biotin operon repressor and
DE   biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15).
GN   BIRA OR RSC0312 OR RS03279.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
```

```
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13840.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   291 AA;  30057 MW;  B9F1FD254D8F18C7 CRC64;

  Query Match             10.5%;  Score 324;  DB 16;  Length 291;
  Best Local Similarity   32.9%;  Pred. No. 5.3e-15;
  Matches   95;  Conservative   40;  Mismatches  122;  Indels   32;  Gaps    8;

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELAR----------IAP 105
              ||| |       : :|  |  :|:   :     |:| :::   |              ||
Db         10 WRLTR-------DRIRPTGIAAGWPVEVVESTGSTNADLMAAVRDAAWPVAAGVPAGTAP 62

Qy        106 DKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRR 165
                  : :     |: ||||||| |    |  | ||    |     :|  ||    ||
Db         63 LVGGRVLAALR-QTAGRGRQGRPWDGDHG--LTFSVACAFAAEPAQLAGLSLAVGVAVAE 119

Qy        166 ALSRL----GLDVQ---IKWPNDLVVGRDKLGGILIETVRTG-GKTVAVVGIGINFVLPK 217
              |::      | : |   :|||||: :   || |||:||:|   |:| ||:|||:|   |
Db        120 AVAAYAQTHGGNAQALALKWPNDVQIAGRKLAGILVETMRAAPGQTWAVIGIGLNLERPH 179

Qy        218 EVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDH 277
              :| ||  :||          :   :|  ||  |   | ::   | ||||| : | :
Db        180 ALE-AALGRSLSGVEELVERPEPNAVLAALLSALGEHLQRFGSHGLAPFVAPFAARDAFA 238

Qy        278 GKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDD 326
              |: | | :||  | ||   |:| || | :|:    | : |||:|||: |
Db        239 GRPVRLWQDGAVVLEGVAHGIDAQGRLAIESGGRVQWIHSGEVSLRAAD 287


RESULT 4
Q8X709
ID   Q8X709      PRELIMINARY;      PRT;   321 AA.
AC   Q8X709;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Biotin-[acetylCoA carboxylase] holoenzyme synthetase and biotin operon
DE   repressor (Bifunctional protein BirA biotin-[acetylCoA carboxylase]
DE   holoenzyme synthetase / biotin operon repressor).
GN   BIRA OR Z5544 OR ECS4900.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005629; AAG59173.1; -.
DR   EMBL; AP002567; BAB38323.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   321 AA;  35329 MW;  E1C1BDDD39E2B68D CRC64;

  Query Match             10.0%;  Score 308.5;  DB 16;  Length 321;
  Best Local Similarity   31.3%;  Pred. No. 7.7e-14;
  Matches  104;  Conservative   51;  Mismatches  144;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |   :|       ||: |   |           ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|| :  ||:  |    :     |:|  :|:  ||   |:    ||   |
Db         59 -SLPEPIQLLNAEQI--LGQLDGGSVTVLPVIDSTNQYLLD--RIGELKSGDA-CVAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQI 176
              ||||:||||    |  |  |  |  :: |   :| || |  :      | :|| | |::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
              | |   |  |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 5
Q9KV39
ID   Q9KV39      PRELIMINARY;      PRT;   320 AA.
AC   Q9KV39;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein.
GN   VC0319.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
```

```
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004120; AAF93492.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; VC0319; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   320 AA;  35040 MW;  88E6C496EBEFB5B6 CRC64;

  Query Match             9.6%;  Score 294.5;  DB 16;  Length 320;
  Best Local Similarity   28.9%;  Pred. No. 7.5e-13;
  Matches   98;  Conservative   45;  Mismatches  145;  Indels   51;  Gaps   11;

Qy         11 VLAELADGLPQHVSQ-------LARMADMKPQQLNGFWQQMPAHIRGLLR--------QH 55
              :| :||||   |  :       ::| |  |            ||:|:         |
Db         10 ILKQLADG-DFHSGEVLGAQLGISRAAISK------------HIQGIRDWGVDVFRVQG 55

Qy         56 DGYWRLVRPLAVFDAEGLRELGERSGFQTALKHE------CASSNDEILELARIAPDKAH 109
               || :| : : : |          :|  |: : :         |:|  :|:
Db         56 KGY-QLAQAMTLLD---------QSVIQSQVNNPVELHPIIGSTNQYLLDHVETL---VS 102

Qy        110 KTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSR 169
              |:|:   |: ||||:|| |    |  |  |  |  |         || :  ||   ||
Db        103 GTVCLAEYQASGRGRRGRHWVSPFGANLYLSIYWRLDAGMAAAMGLSLVVGVAIVEALEA 162

Qy        170 LGLD-VQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQS 227
              :|:| |::|||||     || |||:| : : |     |:|:|||  :     |
Db        163 MGVDGVKLKWPNDLYYQDKKLAGILVEMSGQAGAAAHLVIGMGINLAMRDNEGNIDQPWI 222

Qy        228 LFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDG 287
                  :  |  |  |:  ||  | ||   |  ||  :   :   |: | ||  |
Db        223 SLAEVTGQSRIDRNALAINLIAALDRTLRQYEISGMQNFVERWNRWDNFIGRPVKLLM-G 281

Qy        288 ETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDD 326
                |  :|:|  | : ||| || :: : ||||||  :|
Db        282 ANEVRGIERGIDEHGGVLLETEEGLKSFIGGEISLRKND 320


 ESULT 6
Q9ZIU6
ID   Q9ZIU6      PRELIMINARY;      PRT;   288 AA.
AC   Q9ZIU6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative biotin protein ligase.
GN   BIRA.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016461; AAC68833.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase.
SQ   SEQUENCE   288 AA;  29666 MW;  39F6E9CDBC5F6F77 CRC64;

  Query Match             9.5%;  Score 293;  DB 2;  Length 288;
  Best Local Similarity   32.9%;  Pred. No. 8.2e-13;
  Matches   80;  Conservative   43;  Mismatches  106;  Indels   14;  Gaps    5;

Qy         92 SSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYE 151
              |:| ::|  || |   |   :   |||  |||| || | :|:|  ||||  :  | |  :
Db         46 STNADLLARAR-AGGGARPCLLGAHLQETGRGRAGRPWQNRVGAALMFSCAFDVDLPPSQ 104

Qy        152 LGSLSPVAAVACRRALSRL-----GLDVQIKWPNDLVVGRDKLGGILIETVRTGGKT--V 204
              | :|||:| :|   || :|     ||   :|||||:     || |:|:|:||    |
Db        105 LPALSPLAGMAACEALRQLVGQPAGL--LVKWPNDVQWHDAKLAGVLVESVRNPAGTGHA 162

Qy        205 AVVGIGINF----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYAR 260
               |:|:|:|      | : :: : :  :       :|   || |:  :          :
Db        163 VVIGMGLNLHGGDALSQALQRSVADWNDVAGPGEKGPDAAADLVAAVARAWQDAAHTLQQ 222

Qy        261 DGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEI 320
              :||  | | :   :   |: | ::  || :: |   |||  | | :: |:| : |  |||
Db        223 EGFGAFGARFARVDALAGRPVNVIDRGEILYTGAASGVDDHGRLLVQAADGMRPVSVGEI 282

Qy        321 SLR 323
              |:|
Db        283 SIR 285


RESULT 7
Q8Y2M4
ID   Q8Y2M4      PRELIMINARY;      PRT;   295 AA.
AC   Q8Y2M4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcription regulation accessory factor transcription
DE   regulator protein.
GN   RSC0311 OR RS03278.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13839.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Complete proteome.
SQ   SEQUENCE   295 AA;  30990 MW;  133074A7764BFCBB CRC64;

  Query Match             9.2%;  Score 284.5;  DB 16;  Length 295;
  Best Local Similarity   30.3%;  Pred. No. 3.4e-12;
  Matches   86;  Conservative   49;  Mismatches  102;  Indels   47;  Gaps   11;

Qy        341 LLLDGGNSRLKWAW----------VENGT-FATVGSAPYRDLSPLGAEW------AEKAD 383
              ||:| ||:|:||||           || :   |: |:  |: | :|      |  |
Db         14 LLIDAGNTRIKWAWTAADVAPPAVAPGGTPWQHAGARPHDQLAELVEDWRDCHAGAGMAP 73

Qy        384 GNVRIVGCAVCGEFKKAQVQEQLAR-----KIEWLPSSAQALGIRNHYRHPEEHGSDRWF 438
```

```
                :| |    | |   :  :  ::||     ::  : | | | |:|| || | : |:|||
Db        74 PDVWI--SVVAGPALRDALCARIARVFDGARLRIVASEAAAAGLRNGYRDPAQLGTDRWV 131

Qy       439 NALGSRR-FSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANL--- 494
             |:|:|  :   | ::|: ||| |:| :  || : || |:||  ||  :|:  || |
Db       132 GAVGARHAWPDTALLLVTAGTATTLDIVAPDGRFAGGLILPGLTLMMRALSRNTAQLPEI 191

Qy       495 ---------NRHAGKRYP-FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVI 544
                     :  |    | :  | :|:| | : |  |::      |: :     |   :
Db       192 DIGYLAARDDAQAPADVPSWADNTQDAIALGCVTAQAGAIAQTWQALQAQYPG--PYRCV 249

Qy       545 ITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGRE 588
             ::||  | :|  |   |      :: ||||:.| | ::|  | |
Db       250 LSGGARAALAPHLRMPF------QMHDNLVLLG-LQVLAHTGHE 286


RESULT 8
Q9HWC0
ID   Q9HWC0      PRELIMINARY;      PRT;   312 AA.
AC   Q9HWC0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein.
GN   BIRA OR PA4280.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07668.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   312 AA;  34049 MW;  AC92F3BA43C4BCAF CRC64;

  Query Match             9.1%;  Score 279.5;  DB 16;  Length 312;
  Best Local Similarity   33.1%;  Pred. No. 8.3e-12;
  Matches   90;  Conservative   38;  Mismatches  125;  Indels   19;  Gaps    8;

Qy        57 GYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTH 116
             || ||  ||:: |   :  |    :   :     |:| | :   |:|   :   : |
Db        52 GY-RLASPLSLLDPRKIDSL-----WPVEIMPTVDSTNSEAMR--RLAAGASTPFVIVAE 103

Qy       117 LQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-DVQ 175
             |: ||||:|| |    || | :|          ||  || |  :|  |||  ||: |
Db       104 QQTAGRGRRGRAWVSPYGENLYYSLAITVRGGARELEGLSLVVGLAVLRALESLGVQDTG 163

Qy       176 IKWPNDLVVGRDKLGGILIETVRTGGKTV----AVVGIGINFVLPKEVENAASVQSLFQT 231
             :||||||: |  |: |||:|     |        |:|||||  :    :       || :
Db       164 LKWPNDLLQGGKKIAGILLE---LSGDPADLCHVVIGIGINVNMKTATDIGQPWTSLREA 220

Qy       232 ASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                 | : ||  |  :|   | :::|:|||    |::: |   | | ::|    :
Db       221 LGEL--VDRSRLLSALNRQLADYLDRHSREGFAASREEWESCNLWQG-AQVILSTALSSI 277

Qy       292 EGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
             || | ||| :| | ||    :|    ||:|||

Db       278 EGRVLGVDSRGALRLEVDGVEQGFSGGELSLR 309


RESULT 9
Q8Z317
ID   Q8Z317      PRELIMINARY;      PRT;   320 AA.
AC   Q8Z317;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional protein: biotin operon repressor and biotin-[acetyl-CoA
DE   carboxylase] synthetase.
GN   STY3741.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627279; CAD09497.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   320 AA;  35396 MW;  B584A48BABBEF59F CRC64;

  Query Match             9.0%;  Score 275.5;  DB 16;  Length 320;
  Best Local Similarity   29.9%;  Pred. No. 1.6e-11;
  Matches   98;  Conservative   50;  Mismatches  147;  Indels   33;  Gaps   13;

Qy        11 VLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GYW 59
             ::: ||||      ||     |   :|       ||: |    |          ||
Db        10 LISLLADGEFHSGEQLGERLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY- 58

Qy        60 RLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQS 119
             |  |: : ||: :    :       |:     |:|  :|:  ||   ::    ||  |
Db        59 SLPEPIQLLDADRIH--SQLDSGNVAVLPVIDSTNQYLLD--RIGELRSGDA-CVAEYQQ 113

Qy       120 KGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQIK 177
             ||||:||||     | |  |  |  :: |   :| || |  :     |  || | |::|
Db       114 AGRGRRGRKWFSPFGANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRELGADKVRVK 172

Qy       178 WPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRG 236
             ||||| :   || |||:| | :||     |:| |||  : : ||      |
Db       173 WPNDLYLLDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGI 231

Qy       237 NADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVK 296
               |  :|   |: :| | |  : ::| :|::: ::  :    : | |:   : :| |  :
Db       232 TLDRNMLAAKLIYKLRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLIIGDKEIF-GISR 290

Qy       297 GVDGQGVLHLETAEGKQTVVSGEISLRS 324
             |:| || | ||     :  : ||||||
Db       291 GIDTQGALLLEQDGVIKPWMGGEISLRS 318
```

```
RESULT 10
Q8ZAN5
ID   Q8ZAN5      PRELIMINARY;      PRT;   319 AA.
AC   Q8ZAN5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional protein BirA (EC 6.3.4.15).
GN   BIRA OR YPO3759.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414158; CAC93227.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   319 AA;  35164 MW;  B58E9306939E3B2A CRC64;

  Query Match             8.9%;  Score 274.5;  DB 16;  Length 319;
  Best Local Similarity   29.0%;  Pred. No. 1.9e-11;
  Matches   96;  Conservative   46;  Mismatches  150;  Indels   39;  Gaps   12;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFD 69
              |::: |:||      ||     |   :|   |     ||          |  | ||
Db          9 RLVSILSDGFFHSGEQLGETLGMSRAAINKHMQT----IRD--------W----GLDVFT 52

Qy         70 AEG--------LRELGER-------SGFQTALKHECASSNDEILELARIAPDKAHKTICV 114
                |        :: | |:       :|  | |     |:|  :|:  ||   |:    ||
Db         53 VPGKGYSLPTSIQLLDEQKILSYLPTGHVTVLP-VVDSTNQYLLD--RITELKSGDA-CV 108

Qy        115 THLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLD- 173
                |  ||||:||:|    |  |  |  |  ::       || |  :     |  || :
Db        109 AEYQQAGRGRRGRQWVSPFGANLYLSMYWRLEQGPAAAMGLSLVVGIVMAEVLQNLGAEK 168

Qy        174 VQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTA 232
              |::|||||| :   || |||:| | :||     |:| |||  :     || : |
Db        169 VRVKWPNDLYLNDKKLAGILVELTGKTGDAAQLVIGAGINLTMRGSETNAIN-QDWINLQ 227

Qy        233 SRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFE 292
                  |   |   :| ||   :::::  :| : |:: :|  :    : | |:   : :|
Db        228 DAGVTIDRNKLTAEILSELRLAVVKFENEGLSAFLSRWQKMDNYLNRPVKLIIGNQEIF- 286

Qy        293 GTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              |  :|:| || | ||     :  : ||||||
Db        287 GIARGIDQQGALLLEQDGNIKPYIGGEISLR 317


RESULT 11
O32511
ID   O32511      PRELIMINARY;      PRT;   322 AA.
AC   O32511;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ORF105d (Fragment).
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KD8301;
RX   MEDLINE=20075120; PubMed=10606814;
RA   Narumi I., Satoh K., Kikuchi M., Funayama T., Kitayama S.,
RA   Yanagisawa T., Watanabe H., Yamamoto K.;
RT   "Molecular analysis of the Deinococcus radiodurans recA locus and
RT   identification of a mutation site in a DNA repair-deficient mutant,
RT   rec30.";
RL   Mutat. Res. 435:233-243(1999).
DR   EMBL; AB005471; BAA21331.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   322 AA;  33999 MW;  183BCDA5CBA9C31B CRC64;

  Query Match             7.9%;  Score 244;  DB 2;  Length 322;
  Best Local Similarity   32.5%;  Pred. No. 2.8e-09;
  Matches  111;  Conservative   42;  Mismatches  121;  Indels   68;  Gaps   20;

Qy          9 WRVLAELADGL---PQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPL 65
              ||: | |   |   ||    | |  |:   :|   :::          | ||      |:
Db         15 WRMPARLLPLLTDRPQSGDDLGRALDLGRVTVNTLARRL---------QEDGV-----PV 60

Qy         66 AVFDAEGLRELGE--------RSGFQTALKHE--CASSNDEILELARIAPDK-----AHK 110
               :  |    | |          |  |  ||:::    |: ||:    |   |        |
Db         61 LISRAGYALEPGTPAPGLVPLRGQFGRALRYQGTVTSTQDEV----RFWADSERGPAPHG 116

Qy        111 TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSL--SPVAA-VACRRAL 167
               :  |   |: ||||:|| |    |  | ||   |  |    |  |   |:|| ||   |
Db        117 AVVVAERQTAGRGRRGRVWDTTHG-TLAFS---VLLREALSLADLVRMPLAAGVALHAAC 172

Qy        168 SRLGLDVQIKWPNDLVV--GRDKLGGILIETVRTGGKT-VAVVGIGINFVLPKEVENAAS 224
              : ||:   :||||||:   || || |||:|    | :   ||:|||||  :    | ||
Db        173 TALGVPCGLKWPNDLLTPDGR-KLAGILLEADLRGEEARRAVLGIGIN--VGTAPEGAAH 229

Qy        225 VQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLL 284
              :       :|     |||| | :| :|:  | | |       :| ::| :   |: | :
Db        230 LHESVPDLTR-----AAVLGE-VLFQLEHWLAQPA----PAIIATWKATSLTLGRPVRV- 278

Qy        285 RDGETVFEGTVKGV----DGQGVLHLETAEGKQTVVSGEISL 322
                  |   |||:||    | || | :|| | :|| :|:: |
Db        279 ----TTARGTVEGVARDLDEQGNLLVETAAGLETVSAGDVQL 316


RESULT 12
Q9RRZ1
ID   Q9RRZ1      PRELIMINARY;      PRT;   474 AA.
AC   Q9RRZ1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BIRA bifunctional protein.
GN   DR2341.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE002065; AAF11888.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; DR2341; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   474 AA;  49966 MW;  58E1E537E0D9B0EE CRC64;

  Query Match             7.9%;  Score 244;  DB 16;  Length 474;
  Best Local Similarity   32.5%;  Pred. No. 5e-09;
  Matches  111;  Conservative   42;  Mismatches  121;  Indels   68;  Gaps   20;

Qy          9 WRVLAELADGL---PQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPL 65
              ||: | |    |   ||    | |  |:    :|   :::          | ||      |:
Db        167 WRMPARLLPLLTDRPQSGDDLGRALDLGRVTVNTLARRL---------QEDGV-----PV 212

Qy         66 AVFDAEGLRELGE--------RSGFQTALKHE--CASSNDEILELARIAPDK-----AHK 110
               :  |    | |          |  |  ||:::    |: ||:    |   |       |
Db        213 LISRAGYALEPGTPAPGLVPLRGQFGRALRYQGTVTSTQDEV----RFWADSERGPAPHG 268

Qy        111 TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSL--SPVAA-VACRRAL 167
               : |    |: ||||:|| |    |  | ||   |  |    |  |   |:|| ||   |
Db        269 AVVVAERQTAGRGRRGRVWDTTHG-TLAFS---VLLREALSLADLVRMPLAAGVALHAAC 324

Qy        168 SRLGLDVQIKWPNDLVV--GRDKLGGILIETVRTGGKT-VAVVGIGINFVLPKEVENAAS 224
               : ||:   :||||||:   || || |||:|    | :   ||:|||||  :    | ||
Db        325 TALGVPCGLKWPNDLLTPDGR-KLAGILLEADLRGEEARRAVLGIGIN--VGTAPEGAAH 381

Qy        225 VQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLL 284
               :      :|     |||| | :| :|:  | | |       :| ::| :   |: | :
Db        382 LHESVPDLTR-----AAVLGE-VLFQLEHWLAQPA----PAIIATWKATSLTLGRPVRV- 430

Qy        285 RDGETVFEGTVKGV----DGQGVLHLETAEGKQTVVSGEISL 322
                |   |||:||    | || | :||| | :|| :|:: |
Db        431 ----TTARGTVEGVARDLDEQGNLLVETAAGLETVSAGDVQL 468


RESULT 13
Q9CNX6
ID   Q9CNX6      PRELIMINARY;      PRT;   312 AA.
AC   Q9CNX6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA.
GN   BIRA OR PM0296.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006065; AAK02380.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   312 AA;  35214 MW;  0DD8B737C7C0E29B CRC64;

  Query Match             7.7%;  Score 238;  DB 16;  Length 312;
  Best Local Similarity   28.0%;  Pred. No. 7.1e-09;
  Matches   79;  Conservative   55;  Mismatches  122;  Indels   26;  Gaps   10;

Qy         52 LRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAH-- 109
              : |   ::||:  |:  :   | :    : :   :|    |:|  :|:      : ||
Db         44 IEQQVQHFRLIPQLSRLNESYLTQ--ALAPYPLYIKPIINSTNQFLLD-----NIAHLE 95

Qy        110 -KTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALS 168
                 ||:   |: ||||:||:|       :: |  |  |: :  |  || |  :|    |
Db         96 KGAICLAEYQTAGRGRRGRQWLSPFAGQVIMSLYWTCDQ-KVNLEGLSLVVGMAIAETLK 154

Qy        169 RLG-LDVQIKWPNDLVVGRDKLGGILIETV-RTGGKTVAVVGIGINFVLPKEVENA---- 222
              : | |:: :|||||:::   || |||:|   |   :   |:| |||   ||: :
Db        155 QAGALNIGLKWPNDVLLHGRKLAGILVEIANRQNNQHNLVIGFGINLSFPKQTQQIYQPW 214

Qy        223 ASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGF-APFVAEYQAANRDHGKAV 281
              | :  :  |  |         |:  |  :| | |  : : |    |  ::: :  |: |
Db        215 AELIEVLPTIDRN------KLIAELGKKLIARLQHFEQAGIDEAFRQDWRDMDAYFGEEV 268

Qy        282 LLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
               :| : : :  |  :|:| :| | | |:|| |    ||:|||
Db        269 HVLTENQKI-TGIAQGIDKRGYLQLITSEGIQYFNGGEVSLR 309


RESULT 14
Q9PCI3
ID   Q9PCI3      PRELIMINARY;      PRT;   317 AA.
AC   Q9PCI3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional transcriptional repressor of the biotin operon/biotin
DE   acetyl-CoA-carboxylase synthetase.
GN   XF1796.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
```

```
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004001; AAF84604.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   317 AA;  33715 MW;  DB91D033A7DD07D7 CRC64;

  Query Match             7.7%;  Score 235.5;  DB 16;  Length 317;
  Best Local Similarity   28.6%;  Pred. No. 1.1e-08;
  Matches   83;  Conservative   46;  Mismatches  120;  Indels   41;  Gaps   11;

Qy         53 RQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECA------------SSNDEILE 99
              |   ||  |  ||:: :|| :|           | |  |             |:|  :|
Db         48 RAEQGY-ALAAPLSLLNAERIR---------AELSHSIAAELRVLDVAWTLESTNTSLLA 97

Qy        100 LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVA 159
                :|       : :   |: |||::|| |   |   :  |   ||      :|  ||
Db         98 RPALADG---VEVLLAERQTGGRGQRGRIWISPLAANIYLSLSRVFSSGAAQLMGLSVAV 154

Qy        160 AVACRRALSRLGLD-VQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLP-- 216
               ||   ||   |   |::|||||||    ||||:|:|     |:  ||:|:|:|  :|
Db        155 GVAIAEALIAAGFSHVRLKWPNDLVDNDRKLGGVLVE---RQGEARAVIGLGLNVRMPTN 211

Qy        217 --KEVENA-ASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAA 273
                 ::::     :::|  :   | ||  |:||  ||  |    |    :| |||:  |
Db        212 FAEQIDQPWVDLETLAGSTVER-NALVALLLSALLPALHLFQL----EGLAPFLPRYAKL 266

Qy        274 NRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
               :   | ||  : :    : |  :|:   | | :|     :   :|::|:|
Db        267 DCLVGCAV-WIEEANLKYVGIARGLAADGALLVEINGVLRPFHTGQVSVR 315


RESULT 15
Q9HWC1
ID   Q9HWC1      PRELIMINARY;      PRT;   248 AA.
AC   Q9HWC1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA4279.
GN   PA4279.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07667.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001230; Prenyl_site.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   248 AA;  26758 MW;  609D37480899DA3B CRC64;

  Query Match             7.6%;  Score 234.5;  DB 16;  Length 248;
  Best Local Similarity   27.6%;  Pred. No. 8.9e-09;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Qy        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Qy        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506
              :  |||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Qy        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561
                | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

Qy        562 LAENTVRVADNLVIYGL 578
                    |:  :||  ||
Db        231 ------RIMPDLVFVGL 241


Search completed: June 24, 2003, 21:59:43
Job time : 50.2837 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 10.2965 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-43
Perfect score: 3078
Sequence: 1 MTVLKLSHWRVLAELADGLP.....LVIYGLLNMIAAEGREYEHI 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	307.5	10.0	321	1	BIRA_ECOLI	P06709 escherichia
2	277.5	9.0	320	1	BIRA_SALTY	P37416 salmonella
3	259	8.4	302	1	BIRA_HAEIN	P46363 haemophilus
4	254.5	8.3	267	1	BAF_BORPE	Q45338 bordetella
5	222.5	7.2	325	1	BIRA_BACSU	P42975 bacillus su
6	192.5	6.3	240	1	BIRA_PARDE	P29906 paracoccus
7	133	4.3	233	1	YACB_BACSU	P37564 bacillus su
8	124.5	4.0	726	1	BPL1_HUMAN	P50747 h biotin--p
9	121.5	3.9	1655	1	OMPB_RICCN	Q9kka3 r outer mem
10	116	3.8	1132	1	VHSJ_LAMBD	P03749 bacteriopha
11	111	3.6	752	1	CTPB_MYCTU	Q10877 mycobacteri
12	111	3.6	806	1	HYF2_ALCEU	O07451 alcaligenes
13	110.5	3.6	776	1	HYPF_AZOCH	Q43950 azotobacter
14	109.5	3.6	3172	1	ERY3_SACER	Q03133 saccharopol
15	108	3.5	471	1	GLTD_ECOLI	P09832 escherichia
16	108	3.5	1656	1	OMPB_RICJA	O06653 r outer mem
17	107.5	3.5	573	1	CH60_DROME	O02649 drosophila
18	107	3.5	1524	1	RPOC_THEAQ	Q9kwu6 thermus aqu
19	106.5	3.5	1756	1	TRI2_ECOLI	P22706 escherichia
20	105.5	3.4	1654	1	OMPB_RICRI	Q53047 r outer mem
21	103.5	3.4	690	1	BPL1_YEAST	P48445 s biotin--p
22	103	3.3	356	1	FRVX_ECOLI	P32153 escherichia
23	101.5	3.3	1325	1	MRP4_HUMAN	O15439 homo sapien
24	101.5	3.3	1756	1	TRI1_ECOLI	P14565 escherichia
25	101.5	3.3	3674	1	SPCR_HUMAN	Q9nrc6 homo sapien
26	101	3.3	1448	1	TRC5_ECOLI	P27190 escherichia
27	100.5	3.3	1316	1	RPOC_MYCTU	P47769 mycobacteri
28	100	3.2	1538	1	PPSB_MYCTU	Q10978 mycobacteri
29	98.5	3.2	300	1	RNH3_STRPY	Q99y70 streptococc
30	98	3.2	721	1	ATZN_SYNY3	Q59998 synechocyst
31	97.5	3.2	1059	1	Y379_HUMAN	O15084 homo sapien
32	97	3.2	775	1	PLSB_MYCLE	Q9x7b0 mycobacteri
33	97	3.2	2145	1	U520_CAEEL	Q9u2g0 caenorhabdi
34	96.5	3.1	506	1	ATPA_ANASP	P12405 anabaena sp
35	96.5	3.1	700	1	TKT_MYCTU	O06811 mycobacteri
36	96	3.1	543	1	CH60_MYCGE	P47632 mycoplasma
37	96	3.1	1259	1	YTFN_ECOLI	P39321 escherichia
38	95.5	3.1	488	1	FLAB_PSEAE	P72151 pseudomonas
39	95.5	3.1	544	1	CH60_NEIFL	P48215 neisseria f
40	95.5	3.1	575	1	ILVD_MYCTU	O07433 mycobacteri
41	95.5	3.1	580	1	RUBA_CHLRE	Q42694 chlamydomon
42	95.5	3.1	841	1	RELA_STRAT	O85709 streptomyce
43	95	3.1	732	1	IUTA_ECOLI	P14542 escherichia
44	95	3.1	906	1	HELY_MYCTU	Q10701 mycobacteri
45	94.5	3.1	540	1	CH60_ERWCA	O66220 erwinia car

ALIGNMENTS

RESULT 1
BIRA_ECOLI
ID BIRA_ECOLI STANDARD; PRT; 321 AA.
AC P06709;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE BirA bifunctional protein [Includes: Biotin operon repressor;
DE Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE (Biotin--protein ligase)].
GN BIRA OR BIOR OR DHBB OR B3973.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86006282; PubMed=3899863;
RA Howard P.K., Shaw J., Otsuka A.J.;
RT "Nucleotide sequence of the birA gene encoding the biotin operon
RT repressor and biotin holoenzyme synthetase functions of Escherichia
RT coli.";
RL Gene 35:321-331(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87055242; PubMed=3536662;
RA Buoncristiani M.R., Howard P.K., Otsuka A.J.;
RT "DNA-binding and enzymatic domains of the bifunctional biotin operon
RT repressor (BirA) of Escherichia coli.";
RL Gene 44:255-261(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes.";
RL Nucleic Acids Res. 21:5408-5417(1993).
RN [4]
RP SEQUENCE OF 1-76 FROM N.A.
RC STRAIN=RDD012;
RX MEDLINE=92165747; PubMed=1311302;
RA Pucci M.J., Discotto L.F., Dougherty T.J.;
RT "Cloning and identification of the Escherichia coli murB DNA
RT sequence, which encodes UDP-N-acetylenolpyruvoylglucosamine
RT reductase.";
RL J. Bacteriol. 174:1690-1693(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=93028443; PubMed=1409631;
RA Wilson K.P., Shewchuk L.M., Brennan R.G., Otsuka A.J., Matthews B.W.;
RT "Escherichia coli biotin holoenzyme synthetase/bio repressor crystal
RT structure delineates the biotin- and DNA-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9257-9261(1992).
CC -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS

```
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10123; AAA23520.1; -.
DR   EMBL; M15820; AAA23521.1; -.
DR   EMBL; L14557; AAA24186.1; -.
DR   EMBL; U00006; AAC43075.1; -.
DR   EMBL; AE000471; AAC76951.1; -.
DR   PIR; B24029; BVECBF.
DR   PIR; A24989; A24989.
DR   PDB; 1BIA; 31-OCT-93.
DR   PDB; 1BIB; 31-OCT-93.
DR   EcoGene; EG10123; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   3D-structure; Complete proteome.
FT   DNA_BIND      22     41       H-T-H MOTIF.
FT   VARIANT       61     61       P -> A (IN STRAIN RDD012).
FT   VARIANT       70     70       K -> E (IN STRAIN RDD012).
FT   HELIX          5     14
FT   TURN          15     16
FT   STRAND        20     20
FT   HELIX         22     29
FT   TURN          30     30
FT   HELIX         33     45
FT   TURN          46     47
FT   STRAND        51     53
FT   TURN          54     56
FT   STRAND        57     59
FT   HELIX         69     74
FT   TURN          75     75
FT   STRAND        81     83
FT   STRAND        86     87
FT   HELIX         90    100
FT   TURN         103    104
FT   STRAND       106    109
FT   STRAND       111    112
FT   TURN         127    128
FT   STRAND       130    139
FT   TURN         146    146
FT   HELIX        147    163
FT   TURN         164    165
FT   TURN         167    168
FT   STRAND       170    172
FT   TURN         173    175
FT   STRAND       176    179
FT   TURN         180    181
FT   STRAND       182    192
FT   STRAND       199    208
FT   TURN         226    226
FT   HELIX        227    229
FT   TURN         230    230
FT   HELIX        235    254
FT   TURN         255    256
FT   HELIX        259    268
FT   STRAND       269    269
FT   TURN         272    273
FT   STRAND       275    280
FT   TURN         281    282
FT   STRAND       283    292
FT   TURN         294    295
FT   STRAND       298    302
FT   TURN         303    304
FT   STRAND       305    309
FT   STRAND       313    313
FT   STRAND       316    316
SQ   SEQUENCE   321 AA;  35312 MW;  B80AEBCEEE1BD2D4 CRC64;

  Query Match             10.0%;  Score 307.5;  DB 1;  Length 321;
  Best Local Similarity   31.0%;  Pred. No. 2.9e-15;
  Matches  103;  Conservative   53;  Mismatches  143;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |    :|        ||: |    |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|: :  ||:  |   |:     |:|  :|:  ||   |:    |:  |
Db         59 -SLPEPIQLLNAKQI--LGQLDGGSVAVLPVIDSTNQYLLD--RIGELKSGDA-CIAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQI 176
               ||||:||||    |  |  |  |  :: |   :| || |  :     |.:|| | |::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 2
BIRA_SALTY
ID   BIRA_SALTY     STANDARD;      PRT;   320 AA.
AC   P37416; Q9L9K4;
DT   01-OCT-1994 (Rel. 30, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR STM4138 OR STMF1.2.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
```

```
RL   Nature 413:852-856(2001).
RN   [2]
RP   SEQUENCE OF 1-135 FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=94324848; PubMed=8048842;
RA   Dombrosky P.M., Schmid M.B., Young K.D.;
RT   "Sequence divergence of the murB and rrfB genes from Escherichia coli
RT   and Salmonella typhimurium.";
RL   Arch. Microbiol. 161:501-507(1994).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF170176; AAF33493.1; -.
DR   EMBL; AE008893; AAL22971.1; -.
DR   EMBL; L14816; -; NOT_ANNOTATED_CDS.
DR   HSSP; P06709; 1BIA.
DR   StyGene; SG10028; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
FT   DNA_BIND     22     41       H-T-H MOTIF.
FT   CONFLICT    100    101       EL -> DV (IN REF. 2).
SQ   SEQUENCE   320 AA;  35395 MW;  010E8483093297F5 CRC64;

  Query Match             9.0%;  Score 277.5;  DB 1;  Length 320;
  Best Local Similarity   29.9%;  Pred. No. 4.4e-13;
  Matches   98;  Conservative   51;  Mismatches  146;  Indels   33;  Gaps   13;

Qy         11 VLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GYW 59
              ::: ||||      ||      |    :|        ||: |    |          ||
Db         10 LISLLADGEFHSGEQLGERLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY- 58

Qy         60 RLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQS 119
                |  |: : ||: :    :      |:     |:|  :|:  ||   ::    ||   |
Db         59 SLPEPIQLLDADRIH--SQLDSGNVAVLPVIDSTNQYLLD--RIGELRSGDA-CVAEYQQ 113

Qy        120 KGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQIK 177
              ||||:||||    |  |  |  |  :: |   :| || |  :      | :|| | |::|
Db        114 AGRGRRGRKWFSPFGANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVK 172

Qy        178 WPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRG 236
              ||||| :   || |||:| | :||     |:| |||  : : ||     |
Db        173 WPNDLYLLDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGI 231

Qy        237 NADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVK 296
               |  :|   |: :| | |  : ::| :|::: ::  :    : | |:   : :| |  :
Db        232 TLDRNMLAAKLIYKLRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLIIGDKEIF-GISR 290

Qy        297 GVDGQGVLHLETAEGKQTVVSGEISLRS 324
                  |:| || | ||     :  : |||||||
Db        291 GIDTQGALLLEQDGVIKPWMGGEISLRS 318


RESULT 3
BIRA_HAEIN
ID   BIRA_HAEIN     STANDARD;      PRT;   302 AA.
AC   P46363;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR HI0220.1.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION.
RA   Koonin E.V., Rudd K.E.;
RL   Submitted (SEP-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32708; AAC21889.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; HI0220.1; -.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
SQ   SEQUENCE   302 AA;  34341 MW;  50EE9D6E9820378B CRC64;
```

```
  Query Match             8.4%;  Score 259;  DB 1;  Length 302;
  Best Local Similarity   29.4%;  Pred. No. 9.1e-12;
  Matches   92;  Conservative   56;  Mismatches  131;  Indels   34;  Gaps   13;

Qy         26 LARMADMKPQ---QLNGFWQQMPAHIR-----GLLRQHDGY-WRLVRPLAVFDAEGLREL 76
              |  :|| :|:   :|  | : :   |:     ||    ||  :|||  | : : : :
Db          7 LTYLADCQPKVRSELEKFSKNLEEDIQQLREIGLDILVDGQDYRLVPMLPLLNPQQI--- 63

Qy         77 GERSGFQTALKHE--CASSNDEILE--LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHR 132
                : |  :: ::   :|:|: ||:  |:    |     :||   |: ||||:||:|
Db         64 -STALFPYSIHYQPIISSTNEWILQNILSLKKGD-----LCVAEYQTAGRGRRGRQWLSP 117

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192
               :|||| | || |:  :  || |  :|         |:||:|||||::    |||||
Db        118 FAGQIMFSFYWAFD-PKKSIEGLSLVIGLAIAEV-----LNVQVKWPNDILFDERKLGGI 171

Qy        193 LIETV-RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVEL 251
              |:|      |    |:|||||  | |: | :   |   :      : :   ||  |:  |
Db        172 LVEIANHKNGMLNLVIGIGINVSLSKQTEIS---QPYAEVCEIDPDVERQTLLPKLIQHL 228

 y        252 DAVLLQYARDGF-APFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAE 310
              | : ::|    |   :|: |      : :| : : |  |  :|:| :| | :
Db        229 YTRLNIFEQNGIDEEFQQAWQSYNAFSNSEINVLTE-QGVISGIEQGIDERGYLKVLCGN 287

Qy        311 GKQTVVSGEISLR 323
              |    ||:|||
Db        288 KIQMFNGGEVSLR 300
```

```
RESULT 4
BAF_BORPE
ID   BAF_BORPE      STANDARD;      PRT;   267 AA.
AC   Q45338; Q45373;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Bvg accessory factor.
GN   BAF.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RX   MEDLINE=95325323; PubMed=7601846;
RA   Deshazer D., Wood G.E., Friedman R.L.;
RT   "Identification of a Bordetella pertussis regulatory factor required
RT   for transcription of the pertussis toxin operon in Escherichia
RT   coli.";
RL   J. Bacteriol. 177:3801-3807(1995).
RN   [2]
RP   SEQUENCE OF 1-38 FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 239-267 FROM N.A.
RC   STRAIN=BP536;
RX   MEDLINE=96419162; PubMed=8821935;
RA   Allen A.G., Maskell D.J.;
RT   "The identification, cloning and mutagenesis of a genetic locus
RT   required for lipopolysaccharide biosynthesis in Bordetella
RT   pertussis.";
RL   Mol. Microbiol. 19:37-52(1996).
CC   -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC       A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC       RNA POLYMERASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12020; AAA75361.1; -.
DR   EMBL; AF016461; AAC68834.1; -.
DR   EMBL; X90711; CAA62242.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Transcription regulation; Activator.
SQ   SEQUENCE   267 AA;  27845 MW;  336A615F67B57901 CRC64;
```

```
  Query Match             8.3%;  Score 254.5;  DB 1;  Length 267;
  Best Local Similarity   31.6%;  Pred. No. 1.7e-11;
  Matches   84;  Conservative   44;  Mismatches  113;  Indels   25;  Gaps    9;

Qy        341 LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 395
              :|:| ||||||  | : :   |    ||      ||  ||   |     | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        396 EFKKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 449
              :   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        450 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTG 509
              :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        510 NAVASGMMDAVCGSVM--MMHGRLKEKTGAGKPVDVIITGGG-------AAKVAEALPPA 560
              |:|||:  |  |:::   : || :     |:  :: : |||       | ::      |
Db        183 QAIASGIAAAQAGAIVRQWLAGRQR----YGQAPEIYVAGGGWPEVRQEAERLLAVTGAA 238

Qy        561 FLAENTVRVADNLVIYGLLNMIAAEG 586
              | |       |: |: ||   :||:|
Db        239 FGATPQPTYLDSPVLDGLA-ALAAQG 263
```

```
RESULT 5
BIRA_BACSU
ID   BIRA_BACSU     STANDARD;      PRT;   325 AA.
AC   P42975;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg, and W23 / PY79;
RX   MEDLINE=95247697; PubMed=7730294;
RA   Bower S., Perkins J.P., Yocum R.R., Serror P., Sorokin A.V.,
RA   Rahaim P., Howitt C.L., Prasad N., Ehrlich S.D., Pero J.;
RT   "Cloning and characterization of the Bacillus subtilis birA gene
RT   encoding a repressor of the biotin operon.";
RL   J. Bacteriol. 177:2572-2575(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=96349105; PubMed=8760912;
RA   Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA   Serror P.;
RT   "Sequence analysis of the Bacillus subtilis chromosome region between
RT   the serA and kdg loci cloned in a yeast artificial chromosome.";
RL   Microbiology 142:2005-2016(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38424; AAA92879.1; ALT_SEQ.
DR   EMBL; U20445; AAB60184.1; -.
DR   EMBL; L47709; AAB38447.1; -.
DR   EMBL; Z99115; CAB14160.1; -.
DR   HSSP; P06709; 1BIB.
DR   SubtiList; BG11206; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
FT   DNA_BIND      23     42       H-T-H MOTIF (BY SIMILARITY).
FT   VARIANT       18     18       N -> S (IN STRAIN W23).
FT   VARIANT       89     89       H -> Q (IN STRAIN W23).
FT   VARIANT       92     92       L -> I (IN STRAIN W23).
FT   VARIANT      137    137       V -> I (IN STRAIN W23).
FT   VARIANT      169    169       E -> A (IN STRAIN W23).
FT   VARIANT      219    220       PN -> SD (IN STRAIN W23).
FT   VARIANT      237    237       A -> S (IN STRAIN W23).
FT   VARIANT      311    311       N -> Q (IN STRAIN W23).
FT   MUTAGEN       23     23       G->S: DEREGULATION OF BIOTIN SYNTHESIS.
FT   MUTAGEN       38     38       W->R: DEREGULATION OF BIOTIN SYNTHESIS.
FT   MUTAGEN       58     58       G->E: DEREGULATION OF BIOTIN SYNTHESIS.
SQ   SEQUENCE   325 AA;  36182 MW;  69088BB2CAD09221 CRC64;

  Query Match             7.2%;  Score 222.5;  DB 1;  Length 325;
  Best Local Similarity   27.7%;  Pred. No. 4.6e-09;
  Matches   85;  Conservative   50;  Mismatches  119;  Indels   53;  Gaps   13;

Qy         47 HIRGLLRQ--------HDGYWRLVRPLAVFDAE---GLRE--LGERSGFQTALK------ 87
              ||  | ::          ||  : :|  : ::|   ||:   :|:   :   |
Db         40 HIEELRKEGYEVEAVRRKGYRLIKKPGKLSESEIRFGLKTEVMGQHLIYHDVLSSTQKTA 99

Qy         88 HECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDR 147
              || |::|          ||:    |: |   |: ||||  | |  : |  :  |     |
Db        100 HELANNN---------APE---GTLVVADKQTAGRGRMSRVWHSQEGNGVWMSLILRPDI 147

Qy        148 PQYELGSLSPVAAVACRRALSR-LGLDVQIKWPNDLVVGRDKLGGILIE------TVRTG 200
              |  :   |: :|||| : :    |:  ||||||:::   |  ||| |       ||:
Db        148 PLQKTPQLTLLAAVAVVQGIEEAAGIQTDIKWPNDILINGKKTVGILTEMQAEEDRVRS- 206

Qy        201 GKTVAVVGIGINF-----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVL 255
                ::|||||         | |::: |:  || | |  :   | | ::: :|:  :
Db        207 ----VIIGIGINVNQQPNDFPDELKDIAT--SLSQAAGEK--IDRAGVIQHILLCFEKRY 258

Qy        256 LQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTV 315
              |   || |    :::     |   :  |     | |   |:| :||| ||| || : :
Db        259 RDYMTHGFTPIKLLWESYALGIG-TNMRARTLNGTFYGKALGIDDEGVLLLETNEGIKKI 317

Qy        316 VSGEISL 322
              | :| |
Db        318 YSADIEL 324


RESULT 6
BIRA_PARDE
ID   BIRA_PARDE     STANDARD;      PRT;   240 AA.
AC   P29906;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15) (Biotin--
DE   protein ligase).
GN   BIRA.
OS   Paracoccus denitrificans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Paracoccus.
OX   NCBI_TaxID=266;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13543;
RX   MEDLINE=93136200; PubMed=8422400;
RA   Xu X., Matsuno-Yagi A., Yagi T.;
RT   "DNA sequencing of the seven remaining structural genes of the gene
RT   cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT   of Paracoccus denitrificans.";
RL   Biochemistry 32:968-981(1993).
CC   -!- FUNCTION: ACTIVATES BIOTIN TO FORM BIOTINYL-5'-ADENYLATE AND
CC       TRANSFERS THE BIOTIN MOIETY TO BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L02354; AAA25601.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Biotin; Ligase.
SQ   SEQUENCE   240 AA;  24407 MW;  9FD9184ACC1221D8 CRC64;

  Query Match             6.3%;  Score 192.5;  DB 1;  Length 240;
  Best Local Similarity   31.0%;  Pred. No. 4.8e-07;
  Matches   77;  Conservative   34;  Mismatches  100;  Indels   37;  Gaps   12;

Qy         92 SSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQ-- 149
              |:| | |:|   ||  :     :   |  ||||:||:|    |    |: | :  |||
 b          7 STNAEALKL---APGLSGSAWVLAREQFAGRGRRGREWVMPAGN---FA-GTLVLRPQGG 59

Qy        150 -YELGSLSPVAAVACRRAL-------SRLGLDVQIKWPNDLVVGRDKLGGILIETVRTG- 200
                 || |||:|   ||       :||     ||||||:::   |: |||:|:  :|
Db         60 ALAAAQLSFVAALALYDALGLACGPAARLA----IKWPNDVLLNGGKVAGILLESSGSGP 115

Qy        201 GKTVAVVGIGINFVLPKEVENAASVQ-SLFQTASRRGNADAAVLLETLLVELDAVLLQYA 259
              |    ||||:|        :| :|:       | :|    ||  |     || :   :|
Db        116 GVQAVAVGIGVNLA---GAPDAGAVEPGATPPVSVQGETGHAVDPEEF---LDLLAPAFA 169

Qy        260 R-------DGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312
              |       ||||    : |     |: : : | |     | :|:|  | | |    |:
Db        170 RWQAQLDTYGFAPIRNAWLARAARLGEPI-IARTGTAESHGIFEGIDDSGALILRGPAGR 228

Qy        313 QTVVSGEI 320
              | : : |:
Db        229 QVIPAAEV 236


RESULT 7
YACB_BACSU
ID   YACB_BACSU     STANDARD;      PRT;   233 AA.
AC   P37564;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yacB.
GN   YACB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             4.3%;  Score 133;  DB 1;  Length 233;
  Best Local Similarity   31.7%;  Pred. No. 0.0099;
  Matches   38;  Conservative   19;  Mismatches   49;  Indels   14;  Gaps    2;

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              |:   | :|:| |:||  ||: :     |  :||  ||| |   : ::  |:|| | ||
Db         95 GLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGI 154

Qy        481 HLMKESLAVRTANLNR--------HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLK 532
                 :  |:|  | | | |          ||       | :|: ||::    | |  :  |:|
Db        155 TISTEALYSRAAKLPRIEITRPDNIIGK------NTVSAMQSGILFGYVGQVEGIVKRMK 208


RESULT 8
BPL1_HUMAN
ID   BPL1_HUMAN     STANDARD;      PRT;   726 AA.
AC   P50747; Q99451;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Biotin--protein ligase (EC 6.3.4.-) (Biotin apo-protein ligase)
DE   [Includes: Biotin--[methylmalonyl-CoA-carboxyltransferase] ligase
DE   (EC 6.3.4.9); Biotin--[propionyl-CoA-carboxylase [ATP-hydrolyzing]]
DE   ligase (EC 6.3.4.10) (Holocarboxylase synthetase) (HCS);
DE   Biotin--[methylcrotonoyl-CoA-carboxylase] ligase (EC 6.3.4.11);
DE   Biotin--[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)].
GN   HLCS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
RP   SEQUENCE FROM N.A., AND VARIANT HCS PRO-237.
RC   TISSUE=Liver;
RX   MEDLINE=95144167; PubMed=7842009;
RA   Suzuki Y., Aoki Y., Ishida Y., Chiba Y., Iwamatsu A., Kishino T.,
RA   Niikawa N., Matsubara Y., Narisawa K.;
RT   "Isolation and characterization of mutations in the human
RT   holocarboxylase synthetase cDNA.";
RL   Nat. Genet. 8:122-128(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ohira M., Seki N., Nagase T., Suzuki E., Nomura N., Ohara O.,
RA   Hattori M., Sakaki Y., Eki T., Murakami Y., Saito T., Ichikawa H.,
RA   Ohki M.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Shibuya K., Kudoh J., Minoshima S., Kawasaki K., Nakatoh E.,
RA   Shintani A., Asakawa S., Shimizu N.;
RT   "Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A., AND VARIANTS HCS P-237; W-508; M-550 AND S-581.
RX   PubMed=11735028;
RA   Yang X., Aoki Y., Li X., Sakamoto O., Hiratsuka M., Kure S.,
RA   Taheri S., Christensen E., Inui K., Kubota M., Ohira M., Ohki M.,
RA   Kudoh J., Kawasaki K., Shibuya K., Shintani A., Asakawa S.,
RA   Minoshima S., Shimizu N., Narisawa K., Matsubara Y., Suzuki Y.;
RT   "Structure of human holocarboxylase synthetase gene and mutation
RT   spectrum of holocarboxylase synthetase deficiency.";
RL   Hum. Genet. 109:526-534(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [6]
RP   SEQUENCE OF 1-92 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98163741; PubMed=9503011;
RA   Dahmane N., Ait Ghezala G., Gosset P., Chamoun Z.,
RA   Dufresne-Zacharia M.C., Lopes C., Rabatel N., Gassanova-Maugenre S.,
RA   Chettouh Z., Abramowski V., Fayet E., Yaspo M.L., Korn B.,
RA   Blouin J.L., Lehrach H., Poustka A., Antonarakis S.E., Sinet P.M.,
RA   Creau N., Delabar J.M.;
RT   "Transcriptional map of the 2.5-Mb CBR-ERG region of chromosome 21
RT   involved in down syndrome.";
RL   Genomics 48:12-23(1998).
RN   [7]
RP   VARIANT HCS PRO-237.
RX   MEDLINE=96125714; PubMed=8541348;
RA   Aoki Y., Suzuki Y., Sakamoto O., Li X., Takahashi K., Ohtake A.,
RA   Sakuta R., Ohura T., Miyabayashi S., Narisawa K.;
RT   "Molecular analysis of holocarboxylase synthetase deficiency: a
RT   missense mutation and a single base deletion are predominant in
RT   Japanese patients.";
RL   Biochim. Biophys. Acta 1272:168-174(1995).
RN   [8]
RP   VARIANT HCS ASN-571.
RX   MEDLINE=96414311; PubMed=8817339;
RA   Dupuis L., Leon-Del-Rio A., Leclerc D., Campeau E., Sweetman L.,
RA   Saudubray J.M., Herman G., Gibson K.M., Gravel R.A.;
RT   "Clustering of mutations in the biotin-binding region of
RT   holocarboxylase synthetase in biotin-responsive multiple carboxylase
RT   deficiency.";
RL   Hum. Mol. Genet. 5:1011-1016(1996).
RN   [9]
RP   VARIANTS HCS PRO-237 AND MET-550.
RX   MEDLINE=98057254; PubMed=9396568;
RA   Aoki Y., Suzuki Y., Li X., Sakamoto O., Chikaoka H., Takita S.,
RA   Narisawa K.;
RT   "Characterization of mutant holocarboxylase synthetase (HCS): a Km for
RT   biotin was not elevated in a patient with HCS deficiency.";
RL   Pediatr. Res. 42:849-854(1997).
RN   [10]
RP   VARIANT HCS GLU-333; ILE-462; ASN-571; SER-581 AND THR-610 DEL.
RX   MEDLINE=99204600; PubMed=10190325;
RA   Aoki Y., Li X., Sakamoto O., Hiratsuka M., Akaishi H., Xu L.,
RA   Briones P., Suormala T., Baumgartner E.R., Suzuki Y., Narisawa K.;
RT   "Identification and characterization of mutations in patients with
RT   holocarboxylase synthetase deficiency.";
RL   Hum. Genet. 104:143-148(1999).
CC   -!- FUNCTION: POSTTRANSLATIONAL MODIFICATION OF SPECIFIC PROTEIN BY
CC       ATTACHMENT OF BIOTIN. ACTS ON VARIOUS CARBOXYLASES SUCH AS ACETYL-
CC       COA-CARBOXYLASE, PYRUVATE CARBOXYLASE, PROPIONYL COA CARBOXYLASE,
CC       AND 3-METHYLCROTONYL COA CARBOXYLASE.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[methylmalonyl-CoA:pyruvate
CC       carboxyltransferase] = AMP + diphosphate + [methylmalonyl-
CC       CoA:pyruvate carboxyltransferase].
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[propanoyl-CoA:carbon-
CC       dioxide ligase (ADP-forming)] = AMP + diphosphate + [propanoyl-
CC       CoA:carbon-dioxide ligase (ADP-forming)].
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[3-methylcrotonoyl-
CC       CoA:carbon-dioxide ligase (ADP-forming)] = AMP + diphosphate + [3-
CC       methylcrotonoyl-CoA:carbon-dioxide ligase (ADP-forming)].
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and mitochondrial.
CC   -!- TISSUE SPECIFICITY: MOSTLY EXPRESSED IN MUSCLE, PLACENTA, IN
CC       LESSER EXTENT IN THE BRAIN, KIDNEY, PANCREAS, LIVER AND LUNG.
CC   -!- DISEASE: DEFECTS IN HLCS ARE THE CAUSE OF HOLOCARBOXYLASE
CC       SYNTHETASE DEFICIENCY (HCS) (OR BIOTIN-RESPONSIVE MULTIPLE
CC       CARBOXYLASE DEFICIENCY), AN AUTOSOMAL RECESSIVE DISORDER
CC       CHARACTERIZED BY METABOLIC KETOACIDOSIS, HYPERAMMONEMIA, EXCRETION
CC       OF ABNORMAL ORGANIC ACID METABOLITES, AND DERMATITIS. CLINICAL AND
CC       BIOLOCHEMICAL SYMPTOMS IMPROVE DRAMATICALLY WITH ADMINISTRATION OF
CC       BIOTIN.
CC   -!- SIMILARITY: WITH E.COLI BIRA AND OTHER EUKARYOTIC BIOTIN--PROTEIN
CC       LIGASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D23672; BAA04902.1; -.
DR   EMBL; D87328; BAA13332.1; -.
DR   EMBL; AP000697; BAA89434.1; -.
DR   EMBL; AP000703; BAA89434.1; JOINED.
DR   EMBL; AP000701; BAA89434.1; JOINED.
DR   EMBL; AP000698; BAA89434.1; JOINED.
DR   EMBL; AB063285; BAB68550.1; -.
DR   EMBL; AP001727; BAA95511.1; -.
DR   EMBL; AP001726; BAA95510.1; -.
DR   EMBL; AJ001864; CAA05056.1; -.
DR   Genew; HGNC:4976; HLCS.
DR   MIM; 253270; -.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
```

```
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase; Multifunctional enzyme; Disease mutation.
FT   SIMILAR      563    609       SOME, TO AVIDIN.
FT   VARIANT      237    237       L -> P (IN HCS).
FT                                 /FTId=VAR_005084.
FT   VARIANT      333    333       V -> E (IN HCS; <10% ACTIVITY).
FT                                 /FTId=VAR_009196.
FT   VARIANT      462    462       T -> I (IN HCS; <10% ACTIVITY).
FT                                 /FTId=VAR_009197.
FT   VARIANT      508    508       R -> W (IN HCS).
FT                                 /FTId=VAR_013009.
FT   VARIANT      550    550       V -> M (IN HCS).
FT                                 /FTId=VAR_009198.
FT   VARIANT      571    571       D -> N (IN HCS; ALMOST NO ACTIVITY).
FT                                 /FTId=VAR_009199.
FT   VARIANT      581    581       G -> S (IN HCS; <10% ACTIVITY).
FT                                 /FTId=VAR_009200.
FT   VARIANT      610    610       MISSING (IN HCS; 14% OF ACTIVITY).
FT                                 /FTId=VAR_009201.
FT   CONFLICT     558    558       E -> K (IN REF. 2).
SQ   SEQUENCE   726 AA;  80759 MW;  855B8E52106D675F CRC64;

  Query Match             4.0%;  Score 124.5;  DB 1;  Length 726;
  Best Local Similarity   22.3%;  Pred. No. 0.18;
  Matches   56;  Conservative   47;  Mismatches  101;  Indels   47;  Gaps   12;

Qy        112 ICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRP-QYELGSLSP-------VAAVAC 163
              |:   |::|:|| |  |   :| | : :   :   | : :||   |         || |
Db        496 IVIAARQTEGKGRGGNVWLSPVG-CALSTL--LISIPLRSQLGQRIPFVQHLMSVAVVEA 552

Qy        164 RRALSRL-GLDVQIKWPNDLVVG-RDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVEN 221
               |::     :::::||||||:      |:||:|: :   |     ::| | |       | |
Db        553 VRSIPEYQDINLRVKWPNDIYYSDLMKIGGVLVNSTLMGETFYILIGCGFN------VTN 606

Qy        222 A---ASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHG 278
              :     :  |    :::  |:   |    |:     :|:          : |:|    | |
Db        607 SNPTICINDLITEYNKQHKAELKPLRADYLIARVVTVLE-------KLIKEFQ----DKG 655

Qy        279 KAVLL-------LRDGETVFEG-------TVKGVDGQGVLHLETAEGKQTVVSGEISLRS 324
               :|       : |: |  |         :: |:|  | | :    |:  |  : :
Db        656 PNSVLPLYYRYWVHSGQQVHLGSAEGPKVSIVGLDDSGFLQVHQEGGEVVTVHPDGNSFD 715

Qy        325 DDRPVSVPKRR 335
                | : :||||
Db        716 MLRNLILPKRR 726


ESULT 9
MPB_RICCN
ID   OMPB_RICCN     STANDARD;      PRT;  1655 AA.
AC   Q9KKA3; Q9KK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       SIMILARITY).
CC   -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008659; AAL03623.1; -.
DR   EMBL; AF123721; AAF34124.1; -.
DR   EMBL; AF123726; AAF34129.1; -.
DR   EMBL; AF149110; AAD39533.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall; Complete proteome.
FT   CHAIN          1   1334       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN       1335   1655       32 KDA BETA PEPTIDE.
FT   VARIANT       61     61       P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT       75     75       G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT       78     78       K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      251    251       V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      413    413       N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      959    959       I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      988    988       A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     1139   1139       R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT   CONFLICT     353    354       KD -> GH (IN REF. 3).
FT   CONFLICT     776    776       F -> S (IN REF. 3).
FT   CONFLICT    1159   1159       E -> D (IN REF. 3).
FT   CONFLICT    1177   1177       G -> S (IN REF. 3).
FT   CONFLICT    1492   1492       H -> R (IN REF. 3).
SQ   SEQUENCE   1655 AA;  168342 MW;  E49E19377D5FCE37 CRC64;

  Query Match             3.9%;  Score 121.5;  DB 1;  Length 1655;
  Best Local Similarity   21.1%;  Pred. No. 0.84;
  Matches   99;  Conservative   61;  Mismatches  184;  Indels  125;  Gaps   20;

Qy        181 DLVVGRDKLGGILIETVR---------TGGKTVAVVGIGINFVLPKEVENAASVQSLFQT 231
              |: |  |   | :   |            |||: : | ||  |     :|| :| :
Db        101 DVTVREDTTLGFITNVVNNANHFNLMLNAGKTLTITGQGITNVQAAATKNANNVVAQVNN 160

Qy        232 ASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
               :  | |       :  :|          | |     :  ||    || |:| |   :
Db        161 GAAIDNNDLQG-----VGRIDC--------GAAASTLVFNLANPTTQKAPLILGDNAVIV 207

Qy        292 EG------------------TVKGV---DGQGVLHLETAEGKQTVVSGEISLRSDDRP 328
               |                  ||  :   ||||::    |:   |     ::|:::
Db        208 NGANGTLNVTNGFIKVSSKSFATVNVINIGDGQGIMFNTDADNVNT-----LNLQANGAT 262
```

```
Qy        329 VSVPKRRDSERFLLLD--------------GGNSRLKWAWVENGTFATVGSAPYRDLSPL 374
              ::      : | :||               ||| :     :|  | |  |
Db        263 ITFNGTDGTGRLVLLSKNAAATDFNVTGSLGGNLK---GIIEFNTVAVNG---------- 309

Qy        375 GAEWAEKADGNVRIVG-----------CAVCGEFKKAQVQEQLARKIEWLPSSAQALGIR 423
                :    |  |  ::|              | | :  |:  | : :  || |:|
Db        310 --QLKANAGANAAVIGTNNGAGRAAGFVVSVDNGKVATIDGQVYAK-DMVIQSANAVGQV 366

Qy        424 NHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDA---LTDDGHYLGGTIMPGF 480
              | :||  : |:|      |:  |   |  |     |:| ::    || |:     |:|
Db        367 N-FRHIVDVGTD------GTTAFKTAASKV-----AITQNSNFGTTDFGNLAAQIIVPNT 414

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD---AVCGSVMMMHGRLKEKTGA 537
                :  :     :|    ||            :||   |   ||  ::  :     | :||
Db        415 MTLNGNFTGDASNPGNTAG---VITFDANGTLASASADANVAVTNNITAI-----EASGA 466

Qy        538 GKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEG 586
              |    | ::|  ||::          | :  ::||  || | :|  |  |
Db        467 GV---VQLSGTHAAELRLG-----NAGSVFKLADGTVINGKVNQTALVG 507


RESULT 10
VHSJ_LAMBD
ID   VHSJ_LAMBD     STANDARD;      PRT;  1132 AA.
AC   P03749;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Host specificity protein J.
GN   J.
OS   Bacteriophage lambda.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   Lambda-like viruses.
OX   NCBI_TaxID=10710;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83189071; PubMed=6221115;
RA   Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT   "Nucleotide sequence of bacteriophage lambda DNA.";
RL   J. Mol. Biol. 162:729-773(1982).
CC   -!- FUNCTION: GENE J PROTEIN MAKES UP THE DISTAL FIBER OF THE PHAGE
CC       TAIL. IT IS RESPONSIBLE FOR ADSORPTION TO THE HOST DURING
CC       INFECTION AND DETERMINES HOST SPECIFICITY. IT SPECIFICALLY BINDS
CC       TO THE LAMBDA RECEPTOR PROTEIN OF E.COLI K12.
CC   -!- MISCELLANEOUS: UNDER THE ACTION OF THE GENE G, T, H, M, L, K, AND
CC       I PROTEINS, GENE J PROTEIN SERVES AS THE INITIATOR OF TAIL
CC       POLYMERIZATION. THERE ARE 2-4 COPIES OF PROTEIN J PER MATURE
CC       PHAGE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02459; AAA96553.1; -.
DR   PIR; A04374; QSBPL.
DR   InterPro; IPR003961; FN_III.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 1.
KW   Phage recognition; Fiber protein.
SQ   SEQUENCE   1132 AA;  124422 MW;  53D4AC9044CF3C05 CRC64;

  Query Match             3.8%;  Score 116;  DB 1;  Length 1132;
  Best Local Similarity   19.8%;  Pred. No. 1.3;
  Matches  153;  Conservative   82;  Mismatches  215;  Indels  324;  Gaps   40;

Qy          9 WRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVF 68
              | ||::    :       |   :   |  | |  |:          |  |   |   |
Db        336 WDVLSDFCSAM--------RCMPVWNGQTLTFVQDRPS---------DKTWTYNRSNVVM 378

Qy         69 DAEG-----------------LRELGERSGFQTALKHECASSNDEILE----LARIAPD 106
                :|                    :  :   :|::||          |::|    :||   :
Db        379 PDDGAPFRYSFSALKDRHNAVEVNWIDPNNGWETA---------TELVEDTQAIARYGRN 429

Qy        107 KAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLS-PVAAVACRR 165
                      || : : |   :|:  :|| |         |:      |  ::   | |   |
Db        430 -------VTKMDAFGCTSRGQ--AHRAGL-------WLIKTELLETQTVDFSVGAEGLRH 473

Qy        166 ALSRLGLDVQIKWPNDLV-VGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAAS 224
                          | |:: :  |   ||      |||: :||        | :|:   :|
Db        474 V-----------PGDVIEICDDDYAGI-----STGGRVLAVNSQTRTLTLDREITLPSS 516

Qy        225 VQSLFQTASRRGNADAAVLLETLLVELDAVLLQYAR--DGFAPFVAEYQA---------- 272
                :|       ||    | :|   |  | | :: :|  ||    ||||
Db        517 GTALISLVDGSGN---PVSVEVQSV-TDGVKVKVSRVPDG----VAEYSVWELKLPTLRQ 568

Qy        273 ---------ANRDHGKAVLLLR--------------DGETVFEGTVKGVDGQGVLHLET 308
                      | |   |:  ::                |||    ||| ||    | || |
Db        569 RLFRCVSIRENDDGTYAITAVQHVPEKEAIVDNGAHFDGEQ--SGTVNGVTPPAVQHL-T 625

Qy        309 AEGKQTVVSGE--------------------ISLRSDD---RPVSVPKRRDSE-RFLLL 343
              ||   |  |||                    ::: :||   | ||  :  ::  ||  |
Db        626 AE--VTADSGEYQVLARWDTPKVVKGVSFLLRLTVTADDGSERLVSTARTTETTYRFTQL 683

Qy        344 DGGNSRLKW----AWVENGTFATVG------SAPYR-DLSP------------------ 373
                || ||       || : |  |:|       :|| | :|:|
Db        684 ALGNYRLTVRAVNAWGQQGDPASVSFRIAAPAAPSRIELTPGYFQITATPHLAVYDPTVQ 743

Qy        374 LGAEWAEK-------ADGNVRIVGCAVC------------------------------ 394
                  ::||          : : | :| |:
Db        744 FEFWFSEKQIADIRQVETSTRYLGTALYWIAASINIKPGHDYYFYIRSVNTVGKSAFVEA 803

Qy        395 ---------------GEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWF 438
                            |:  :: : ::|  |:|    :|  |     :    :  ||:|
Db        804 VGRASDDAEGYLDFFKGKITESHLGKELLEKVELTEDNASRL---EEFSKEWKDASDKW- 859

Qy        439 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKES----------LA 488
              ||:                 || ::   |  ||: |  :     |:::           |
Db        860 NAM---------------WAVKIEQTKDGKHYVAGIGLS----MEDTEEGKLSQFLVAA 899

Qy        489 VRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGG 548
              | | ::   |   |    ||                : |    ||  |          || |
Db        900 NRIAFIDPANGNETPMFVAQGN-----------QIFMNDVFLKRLTAP------TITSG 941

Qy        549 GAAKVAEALPPAF--------LAEN--------------TVRVADNLVIYGLL 579
              |        ||||         |:|                | :|:|  | |  |
Db        942 GN-------PPAFSLTPDGKLTAKNADISGSVNANSGTLSNVTIAENCTINGTL 988


RESULT 11
CTPB_MYCTU
ID   CTPB_MYCTU     STANDARD;      PRT;   752 AA.
AC   Q10877; P77905;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Cation-transporting P-type ATPase B (EC 3.6.3.-).
GN   CTPB OR RV0103C OR MT0112 OR MTCY251.22C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RA   Agranoff D.D., Krishna S.K., Griffin G.E., Mangan J.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
```

```
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
CC       (E1-E2 ATPASES). SUBFAMILY IB.
CC   -!- SIMILARITY: CONTAINS 1 HMA DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Y07638; CAA68915.1; -.
DR   EMBL; Z74410; CAA98939.1; -.
DR   EMBL; AE006923; AAK44334.1; -.
DR   HSSP; P04129; 1AFJ.
DR   TIGR; MT0112; -.
DR   TubercuList; Rv0103c; -.
DR   InterPro; IPR001757; ATPase_E1-E2.
DR   InterPro; IPR000579; Cat_ATPaseA_P.
DR   InterPro; IPR001934; HeavyMe_transpt.
DR   InterPro; IPR001454; Hlgnase/hydrlase.
DR   Pfam; PF00122; E1-E2_ATPase; 1.
DR   Pfam; PF00403; HMA; 1.
DR   Pfam; PF00702; Hydrolase; 1.
DR   PRINTS; PR00940; CATPATPASEA.
DR   PROSITE; PS00154; ATPASE_E1_E2; 1.
DR   PROSITE; PS01047; HMA_1; 1.
DR   PROSITE; PS50846; HMA_2; 1.
KW   Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;
KW   Complete proteome.
FT   TRANSMEM    105    125       POTENTIAL.
FT   TRANSMEM    132    152       POTENTIAL.
FT   TRANSMEM    167    187       POTENTIAL.
FT   TRANSMEM    201    221       POTENTIAL.
FT   TRANSMEM    361    381       POTENTIAL.
FT   TRANSMEM    390    410       POTENTIAL.
FT   TRANSMEM    491    511       POTENTIAL.
FT   TRANSMEM    714    734       POTENTIAL.
FT   DOMAIN       16     79       HMA.
FT   MOD_RES     446    446       PHOSPHORYLATION (BY SIMILARITY).
FT   METAL        26     26       BY SIMILARITY.
FT   METAL        29     29       BY SIMILARITY.
FT   CONFLICT     22     22       L -> S (IN REF. 3).
SQ   SEQUENCE   752 AA;  77509 MW;  21563FE4B4407505 CRC64;

  Query Match             3.6%;  Score 111;  DB 1;  Length 752;
  Best Local Similarity   21.7%;  Pred. No. 1.8;
  Matches  146;  Conservative   60;  Mismatches  215;  Indels  252;  Gaps   33;

Qy        31 DMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHEC 90
             |  |:: :| |       | :|     |  :   : ||  |          |:   | :
Db       189 DQPPREGSGIW-------RAILNSDSIYLEVAAGVTVFVLAG-------RYFEARAKSKA 234

Qy        91 ASSNDEILEL-----ARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVF 145
              |:   : ||      | : || |   |  : |: :      :::  | || :  : | |
Db       235 GSALRALAELGAKNVAVLLPDGAELVIPASELKKR------QRFVTRPGETIA-ADGVVV 287

Qy       146 DRP-----QYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKL--GGILIETVR 198
             |           |   || |                    :|   |||   :  | ::||
Db       288 DGSAAIDMSAMTGEAKPVRA----------------YPAASVVGGTVVMDGRLVIEATA 330

Qy       199 TGGKT--VAVVGIGINFVLPKEVENAAS----------------VQSLFQTASRRG---- 236
               |   |:| :         || | :                 |  :|  |   |
Db       331 VGADTQFAAMVRL---------VEQAQTQKARAQRLADHIAGVFVPVVFVIAGLAGAAWL 381

Qy       237 ----NADAA--VLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETV 290
                 || |  | |  |::     |       | |   |   |: |
Db       382 VSGAGADRAFSVTLGVLVIACPCAL------GLATPTAMMVASGR-------------- 420

Qy       291 FEGTVKGVDGQGVLHLETAEGKQTVV--------SGEISLRSDDRPVSVPKRRDSERFLL 342
               |   |:  :|   |||      |||          |:::: :      |    || |  |
Db       421 --GAQLGIFIKGYRALETIRSIDTVVFDKTGTLTVGQLAVSTVTMAGSGTSERDREEVLG 478

Qy       343 LDGGNSRLKWAWVENGT-----FATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEF 397
             |          | ||: :       | | ::|  |  |:         :| | : || | ||
Db       479 LA--------AAVESASEHAMAAAIVAASP--DPGPV--------NGFVAVAGCGVSGE- 519

Qy       398 KKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFN--------ALGSRRF--- 446
                                    :|      |  | |   |           || | |
Db       520 ----------------------VG-----GHHVEVGKPSWITRTTPCHDAALVSARLDGE 552

Qy       447 SRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPT 506
             ||   ||       |:|| :       :  |        :|:| |   | | |  | |    |
Db       553 SRGETVVF-----VSVDGVVRAALTIADT-------LKDSAAAAVAAL-RSRGLRTILLT 599

Qy       507 TTGNAVASGMMDAVCGSV--------MMMHG------RLKEK------------------ 534
               | |    |||   |          |:  |      ||:|:
Db       600 GDNRAAA----DAVAAQVGIDSAVADMLPEGKVDVIQRLREEGHTVAMVGDGINDGPALV 655

Qy       535 -------TGAGKPV-----DVIITGGGAAKVAEALPPAFLAENTVRV----------AD 571
                    | |  |     |:|:        | :||  |     |:|:           |
Db       656 GADLGLAIGRGTDVALGAADIILVRDDLNTVPQALDLARATMRTIRMNMIWAFGYNVAAI 715

Qy       572 NLVIYGLLNMIAA 584
               :   |||| : |
Db       716 PIAAAGLLNPLIA 728


RESULT 12
HYF2_ALCEU
ID   HYF2_ALCEU     STANDARD;      PRT;   806 AA.
AC   O07451;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hydrogenase maturation protein hypF2.
GN   HYPF2.
OS   Alcaligenes eutrophus (Ralstonia eutropha).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=510;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H16 / DSM 428 / ATCC 17699;
RX   MEDLINE=99015922; PubMed=9799289;
RA   Wolf I., Buhrke T., Dernedde J., Pohlmann A., Friedrich B.;
RT   "Duplication of hyp genes involved in maturation of [NiFe]
RT   hydrogenases in Alcaligenes eutrophus H16.";
RL   Arch. Microbiol. 170:451-459(1998).
CC   -!- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS.
CC   -!- SIMILARITY: BELONGS TO THE HYPF FAMILY.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
```

```
CC       ACYLPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF002655; AAB60890.1; -.
DR   InterPro; IPR001792; Acylphosphatase.
DR   InterPro; IPR004421; HypF.
DR   InterPro; IPR000666; Sua5_yciO_yrdC.
DR   Pfam; PF00708; Acylphosphatase; 1.
DR   Pfam; PF01300; Sua5_yciO_yrdC; 1.
DR   ProDom; PD001884; Acylphosphatase; 1.
DR   ProDom; PD002209; Sua5_yciO_yrdC; 1.
DR   TIGRFAMs; TIGR00143; hypF; 1.
DR   PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
KW   Zinc-finger.
FT   DOMAIN        1    101       ACYLPHOSPHATASE-LIKE.
FT   ZN_FING     121    146       C4-TYPE (POTENTIAL).
FT   ZN_FING     171    196       C4-TYPE (POTENTIAL).
SQ   SEQUENCE   806 AA;  86750 MW;  24C32B21FB0E0EBC CRC64;

  Query Match             3.6%;  Score 111;  DB 1;  Length 806;
  Best Local Similarity   20.5%;  Pred. No. 1.9;
  Matches  147;  Conservative   73;  Mismatches  226;  Indels  272;  Gaps   37;

Qy         64 PLAVFDAEGLRELGERSGFQTALKHECASSND----EILELAR--------IAPDKAHKT 111
              |||  |     |:||          ||:  |     ||| :|        |  | |
Db         77 PLARVD-----EIGE---------ERCAAQVDADGFAILESSRSDAAVHTAIGHDTAVCP 122

Qy        112 ICVTHLQSKGRGRQGRKWSHRLGECLM----FSFGWV--FDR------------------ 147
              |:  |       |:: :    |      ::  |   :||
Db        123 DCLAELFDPA----NRRYRYAFINCTQCGPRYTLTWALPYDRATTSMAPFPQCRPCLDEY 178

Qy        148 --PQYELGSLSPVAAVACRRALSRL---GLDVQIKWPNDLVVGRDKLGGILIETVR--TG 200
                |::      | |   |  :|: |   |: |:   |             : |||
Db        179 NDPEHRRFHAEPNACPDCGPSLALLNAQGMPVEDVDP------------IAETVARLQR 225

Qy        201 GKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR-------------------GNADA 240
              |: ||: |:|  | |  :  || :|  |     |                    |: :|
Db        226 GEIVAIKGLG-GFHLACDAHNADAVARLRSRKQREEKPFAVMVANLATAAQWGDIGSGEA 284

Qy        241 AVLLETLLVELDAVLL-QYARDG----------------FAP--FVAEYQAANRDHG-- 278
              |:|  :   |   ||| ::  ||                : |  ::  ::|| |  |
Db        285 ALLTAS---ERPIVLLRKHGVDGRFAGVAPGLVWLGVMLPYTPLQYLLFHEAAGRPEGLG 341

Qy        279 ------KAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSG------EISLRSDD 326
                    ||::       |  | | |      |||:     :          || :| ||
Db        342 WLAQPQSLVLVMTSANPGGEPLVTGND-------ETAQRLTGIADAFLLHDREILVRCDD 394

Qy        327 ----------------------RPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVG 364
                                   | ::   |     | | |          :|    | |
Db        395 SVVRGDGEPAPHVQFIRRARGYTPRAIKLARSGPSVLALGGS--------FKNTVCLTRG 446

Qy        365 SAPY--RDLSPLG------------AEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKI 410
                :  : :  ||             |       :  ::|   :  :|   :   :||
Db        447 DEAFVSQHVGDLGNAATCEALIEAVAHLQRVLEIRPQLVAHDLHPDFFSTRHAAELA--A 504

Qy        411 EWLPSSAQALGIRNHYRH----PEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALT 466
              :|      |: :::|: |      |||||                 : | |:    |
Db        505 QW---GVPAVAVQHHHAHIAAVLAEHGSDE----------------PAIGLALDGVGLG 544

Qy        467 DDGHYLGGTIMPGFH---LMKESLA-------------VRTANLNRHAGKRYP-----F 504
              |||   ||    |      |   |||                | |    ||  |       |
Db        545 DDGQAWGGETAAGGRRRLLAPRSLARLPLPGGDRAAREPWRMAAAALHAMGRGEEIEGRF 604

Qy        505 PTTTG------------NAVASGMM----DAVCG------------SVMMMHGRLK--- 532
              |   |            ||  | |    ||  |            : |:: |  :
Db        605 PRQPGAPMVNRMLAQRLNAPLSSSMGRWFDAAAGLLGTRETMAYEGQAAMLLEGLAESWG 664
```

```
Qy        533 EKTGAGKPVDVIITGGGAAK-----VAEALPPAFLAENTVRVADNLVIYGLLNMIAAE 585
              |:   |:|  |  : ||  :       | ||| |:      :   : | :  ||  ::||
Db        665 EQPSPGRPKTVAHSLGGVPRRGGTYKALALPDAW----RIDAGNTLDLLPLLEALSAE 718


RESULT 13
HYPF_AZOCH
ID   HYPF_AZOCH     STANDARD;      PRT;   776 AA.
AC   Q43950;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hydrogenase maturation protein hypF.
GN   HYPF OR HUPY.
OS   Azotobacter chroococcum mcd 1.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Azotobacter.
OX   NCBI_TaxID=355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252280; PubMed=8486288;
RA   Tibelius K.H., Du L., Tito D., Stejskal F.;
RT   "The Azotobacter chroococcum hydrogenase gene cluster: sequences and
RT   genetic analysis of four accessory genes, hupA, hupB, hupY and
RT   hupC.";
RL   Gene 127:53-61(1993).
RN   [2]
RP   ERRATUM.
RA   Tibelius K.H., Du L., Tito D., Stejskal F.;
RL   Gene 134:141-141(1993).
CC   -!- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HYPF FAMILY.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC       ACYLPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00674; AAA22134.1; ALT_INIT.
DR   HSSP; P41500; 2ACY.
DR   InterPro; IPR001792; Acylphosphatase.
DR   InterPro; IPR004421; HypF.
DR   InterPro; IPR000666; Sua5_yciO_yrdC.
DR   Pfam; PF00708; Acylphosphatase; 1.
DR   Pfam; PF01300; Sua5_yciO_yrdC; 1.
DR   ProDom; PD001884; Acylphosphatase; 1.
DR   ProDom; PD002209; Sua5_yciO_yrdC; 1.
DR   TIGRFAMs; TIGR00143; hypF; 1.
DR   PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
KW   Zinc-finger.
FT   DOMAIN        1     93       ACYLPHOSPHATASE-LIKE.
FT   ZN_FING     113    138       C4-TYPE (POTENTIAL).
FT   ZN_FING     163    188       C4-TYPE (POTENTIAL).
SQ   SEQUENCE   776 AA;  82568 MW;  6ACDA50B8F90911B CRC64;

  Query Match             3.6%;  Score 110.5;  DB 1;  Length 776;
  Best Local Similarity   22.7%;  Pred. No. 2;
  Matches  156;  Conservative   61;  Mismatches  238;  Indels  233;  Gaps   40;

Qy         34 PQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDA---EGLRELGERSGFQTALKHEC 90
              |  |  | :::|| :               |||  ||   | |      :||
Db         54 PAALAEFERRLPAEL-------------PPLARLDALQRENLPAGAHFAGF------AI 93

Qy         91 ASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECL----MFSF--GWV 144
                 | ||       || |    |: |       |:| |   |      :|
Db         94 VPSRDEAGGALAPPPDAAVCPACLEELFDP----HDRRWRHPFINCTHCGPRYSLIRRLP 149
```

```
Qy        145 FDRPQYELGSLS--------------------PVAAVACRRALSRLGLDVQIKWPNDLVV 184
              :|| |  |   :                    |:|  ||   |          |  |
Db        150 YDRAQTSLAGFALCPDCRHEYEDPADRRFHAQPIACPACGPRL----------WCED--A 197

Qy        185 GRDKLGG----ILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR----- 235
                ::| |    : : |:|  |: ||: |:|  | |  :  || :| :| :   ||
Db        198 AGNRLPGEPVELALATLRR-GEIVALRGVG-GFHLACDARNAGAV-ALLRRRKRRPAKPF 254

Qy        236 -------GNADAAV-LLETLLVELDA-----VLLQY--ARDGF-----APFVA------- 268
                    : :| | | || | || |     |||:   | ||      || :|
Db        255 ALMAANPASLEALVELNETGLNELSAPAAPVVLLRRRPAADGLLAAEVAPDLAWLGVMLP 314

Qy        269 --------EYQAANRDHGKA--------VLLLR----DGETVFEGTVK------GVDGQG 302
                      ::|| |  | |         ||::      || :  | ::      |:
Db        315 HSPLHWLLFHEAAGRPAGTAWTAAPQDLVLVMTSANLSGEPLITGNLEAREKLAGITDLW 374

Qy        303 VLH-----------LETAEGKQTVVSGEISLRSDD--RPVSVPKRRDSERFLLLDGGNSR 349
              :||           |  | |:  ||     :||     |  :   |     | | |
Db        375 LLHDREILNRVDDSLVNALGRAPVV-----IRSGRGLAPQEIRLARSGPSVLALGG---- 425

Qy        350 LKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARK 409
                  ::|    | |   :  |||                      |:   |     | |
Db        426 ----QLKNAICLTQGDRAW--LSP------------------HIGDLHSADACRSLERT 460

Qy        410 IEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRN------------------A 450
              :       : ||:|     ||    |   :   | ||:|                    |
Db        461 V---AQFGELLGVR-----PERIACDLHPDFFAS-RFAREHAERHGLPLYPVQHHHAHIA 511

Qy        451 CVVVSCGTAVTVDALTDDGHYLG--GTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTT 508
               |:   | |  |  |  ||  ||  | :  |     | | |          |:  | |
Db        512 AVLAEHGLAEPVLGLALDGFGLGADGRLRGG-----ELLKVAVDGC-AWLGELNPLPLPG 565

Qy        509 GNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL------PPAFL 562
              |:  |        |::  : ||       |:          |:| :|  |      ||:
Db        566 GDQAAREPWRMAAGALHAL-GR------GGEIATRFAAEPGSAAIARMLERGFNCPPSSS 618

Qy        563 AENTVRVADNLVIYGLLNMIAAEGREYE 590
              |      |      ||| :   | : ||
Db        619 AGRLFDAA-----AGLLGL--GERQRYE 639
```

```
RESULT 14
ERY3_SACER
ID   ERY3_SACER     STANDARD;      PRT;  3172 AA.
AC   Q03133; Q54097; Q99270;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
DE   deoxyerythronolide B synthase III) (DEBS 3).
GN   ERYA.
OS   Saccharopolyspora erythraea (Streptomyces erythraeus).
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Saccharopolyspora.
OX   NCBI_TaxID=1836;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=91043075; PubMed=2234082;
RA   Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
RT   "An unusually large multifunctional polypeptide in the erythromycin-
RT   producing polyketide synthase of Saccharopolyspora erythraea.";
RL   Nature 348:176-178(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91220065; PubMed=2024119;
RA   Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
RT   "Modular organization of genes required for complex polyketide
RT   biosynthesis.";
RL   Science 252:675-679(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NRRL 2338;
RX   MEDLINE=92155230; PubMed=1740151;
RA   Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
RT   "6-Deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
RT   Cloning of the structural gene, sequence analysis and inferred domain
RT   structure of the multifunctional enzyme.";
RL   Eur. J. Biochem. 204:39-49(1992).
CC   -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
CC       deoxyerythronolide B.
CC   -!- COFACTOR: NADP; CONTAINS 2 COVALENTLY BOUND PHOSPHOPANTETHEINES.
CC   -!- PATHWAY: COMPLEX POLYKETIDE FORMATION IN ERYTHROMYCIN
CC       BIOSYNTHESIS.
CC   -!- MISCELLANEOUS: IN EACH ORF OF ERYA TWO MODULES ARE PRESENT EACH
CC       ENCODING FOR A FUNCTIONAL SYNTHASE SUBUNIT. THUS ERYA SHOWING 3
CC       ORFS CODES FOR 6 SYNTHASE SUBUNITS. IT IS SUPPOSED THAT EACH
CC       SYNTHASE PARTICIPATES IN ONE OF THE SIX FAS-LIKE ELONGATION STEPS
CC       REQUIRED FOR FORMATION OF THE POLYKETIDE. MODULE 1, 2, 3, 4, 5,
CC       AND 6 PARTICIPATING IN BIOSYNTHESIS STEPS 1, 2, 3, 4, 5, AND 6,
CC       RESPECTIVELY.
CC   -!- MISCELLANEOUS: BIOSYNTHESIS OF POLYKETIDES; ACYLTRANSFERASE (AT),
CC       BETA-KETOACYL CARRIER PROTEIN SYNTHASE (KS), AND ACYL CARRIER
CC       PROTEIN (ACP) FOR CHAIN ELONGATION. BETA-KETOREDUCTASE (KR),
CC       DEHYDRATASE (DH), AND ENOYL REDUCTASE (ER) FOR PROCESSING OF THE
CC       BETA CARBON, AND THIOESTERASE (TE) FOR RELEASE AND LACTONIZATION
CC       OF THE FULL-LENGTH CHAIN.
CC   -!- SIMILARITY: TO FATTY ACID SYNTHASE (FAS).
CC   -!- SIMILARITY: CONTAINS 2 ACYL CARRIER DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X56107; CAA39583.1; -.
DR   EMBL; M63677; AAA26495.1; -.
DR   EMBL; X62569; CAA44449.1; -.
DR   HSSP; P00101; 1CCH.
DR   InterPro; IPR002198; ADH_short.
DR   InterPro; IPR001227; Ac_transferase.
DR   InterPro; IPR000794; Ketoacyl-synt.
DR   InterPro; IPR003880; Ppantne_attach.
DR   InterPro; IPR001031; Thioesterase.
DR   Pfam; PF00106; adh_short; 1.
DR   Pfam; PF00109; ketoacyl-synt; 2.
DR   Pfam; PF00550; pp-binding; 2.
DR   Pfam; PF00698; Acyl_transf; 2.
DR   Pfam; PF00975; Thioesterase; 1.
DR   Pfam; PF02801; ketoacyl-synt_C; 2.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; 2.
DR   PROSITE; PS00606; B_KETOACYL_SYNTHASE; 2.
DR   PROSITE; PS50075; ACP_DOMAIN; 2.
KW   Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Repeat;
KW   Phosphopantetheine; Multifunctional enzyme.
FT   DOMAIN        1   1484       MODULE 5.
FT   DOMAIN     1485   3172       MODULE 6.
FT   DOMAIN       37    484       BETA-KETOACYL SYNTHASE 1.
FT   DOMAIN      554    878       ACYLTRANSFERASE (AT) 1.
FT   DOMAIN     1116   1298       BETA-KETOACYL REDUCTASE 1.
FT   DOMAIN     1394   1464       ACYL CARRIER (ACP) 1.
FT   DOMAIN     1488   1954       BETA-KETOACYL SYNTHASE 2.
FT   DOMAIN     2021   2335       ACYLTRANSFERASE (AT) 2.
FT   DOMAIN     2555   2735       BETA-KETOACYL REDUCTASE 2.
FT   DOMAIN     2821   2891       ACYL CARRIER (ACP) 2.
FT   DOMAIN     2926   3172       THIOESTERASE.
FT   ACT_SITE    199    199       THIOESTER BOND.
FT   ACT_SITE    643    643       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND    1118   1164       NADP.
FT   BINDING    1427   1427       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   ACT_SITE   1661   1661       THIOESTER BOND.
```

```
FT   ACT_SITE    2112   2112       ACYL-ENZYME INTERMEDIATE.
FT   NP_BIND     2557   2605       NADP.
FT   BINDING     2854   2854       PHOSPHOPANTETHEINE (BY SIMILARITY).
FT   CONFLICT     231    231       S -> SA (IN REF. 2).
FT   CONFLICT     240    240       MISSING (IN REF. 2).
FT   CONFLICT     289    289       A -> R (IN REF. 2).
FT   CONFLICT     493    493       P -> R (IN REF. 2).
FT   CONFLICT     493    517       PEPRNSLRDTGFTLATRASAMEHRA -> ASRGTRCATPVS
FT                                 RWPPAAAPWEQ (IN REF. 1).
FT   CONFLICT     510    510       A -> R (IN REF. 2).
FT   CONFLICT     513    513       M -> W (IN REF. 2).
FT   CONFLICT     525    525       E -> D (IN REF. 2).
FT   CONFLICT     536    536       R -> G (IN REF. 2).
FT   CONFLICT     547    551       GPNSP -> ARTR (IN REF. 2).
FT   CONFLICT     553    553       R -> G (IN REF. 2).
FT   CONFLICT     673    673       R -> A (IN REF. 2).
FT   CONFLICT     716    716       MISSING (IN REF. 2).
FT   CONFLICT     734    736       AHK -> GIT (IN REF. 2).
FT   CONFLICT     896    896       R -> RQR (IN REF. 2).
FT   CONFLICT     896    896       R -> RELPVYPFQRQR (IN REF. 1).
FT   CONFLICT     988    994       GVAAVPH -> VSLLSRD (IN REF. 2).
FT   CONFLICT    1108   1116       RTHPLEPLA -> ARTRWSPR (IN REF. 2).
FT   CONFLICT    1124   1126       MISSING (IN REF. 1).
FT   CONFLICT    1132   1132       L -> V (IN REF. 2).
FT   CONFLICT    1192   1192       A -> R (IN REF. 2).
FT   CONFLICT    1194   1194       MISSING (IN REF. 2).
FT   CONFLICT    1277   1278       AA -> RR (IN REF. 2).
FT   CONFLICT    1385   1390       LCDGRE -> STAER (IN REF. 2).
FT   CONFLICT    1485   1485       MISSING (IN REF. 2).
FT   CONFLICT    1518   1518       G -> R (IN REF. 2).
FT   CONFLICT    1601   1601       V -> L (IN REF. 2).
FT   CONFLICT    1724   1725       LP -> FA (IN REF. 2).
FT   CONFLICT    1732   1732       Q -> L (IN REF. 2).
FT   CONFLICT    1739   1743       GPAEG -> ARRA (IN REF. 2).
FT   CONFLICT    1762   1762       T -> S (IN REF. 2).
FT   CONFLICT    2252   2252       D -> DGAD (IN REF. 2).
FT   CONFLICT    2275   2277       QSP -> AVA (IN REF. 2).
FT   CONFLICT    2408   2408       G -> GR (IN REF. 2).
FT   CONFLICT    2420   2421       LA -> S (IN REF. 2).
FT   CONFLICT    2443   2444       NA -> TH (IN REF. 2).
FT   CONFLICT    2596   2596       A -> G (IN REF. 2).
FT   CONFLICT    2609   2609       P -> A (IN REF. 2).
FT   CONFLICT    2715   2722       RRAEGRAA -> AVRKAVRR (IN REF. 1).
FT   CONFLICT    2754   2754       D -> E (IN REF. 2).
SQ   SEQUENCE   3172 AA;  331474 MW;  DBBD5094E77DDD5F CRC64;

  Query Match             3.6%;  Score 109.5;  DB 1;  Length 3172;
  Best Local Similarity   22.5%;  Pred. No. 14;
  Matches  108;  Conservative   59;  Mismatches  205;  Indels  107;  Gaps   24;

Qy         46 AHIRGLLR----QHDGYWRLVRPLAVFDAEGLRELG-----ERSGFQTALKHECASSNDE 96
              | : || |    :  | |  :  |   ||  :| |       |   | |:: :
Db       1041 AMVGGLGRVVGIEQPGRWGGLVDLVDADAASIRSLAAVLADPRGEEQVAIRADG------ 1094

Qy         97 ILELARIAPDKAH------KTICVTHLQSKGRGRQG---RKWSHRLGECLMFSFGWVFDR 147
               :::||: |  |       : :  | | : | |  |    :|  | |   :   |
Db       1095 -IKVARLVPAPARARTHPLEPLAGTVLVTGGTGGIGAHLARWLARSGAEHLVLLG----- 1148

Qy        148 PQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV 207
               |: :|  |   |  |: ||  | |   :  |  | :|  :|       |:||: |
Db       1149 ---RRGADAP-GASELREELTALGTGVTIAACD--VADRARLEAVLAAEAAAEGRTVSAV 1202

Qy        208 --GIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGF-- 263
                 |::  | :   |   :     |       | |: |  :||| :|  :  |
Db       1203 MHAAGVSTSTPLDDLTEAEFTEIADVKVR-----GTVNLDELCPDLDAFVLFSSNAGVWG 1257

Qy        264 APFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEG---KQTVVSGEI 320
              :| :| | |||                :|  :    :|      | |    | :   |
Db       1258 SPGLASYAAAN-------------AFLDGFARAARSEGAPVTSIAWGLWAGQNMAGDEG 1303

Qy        321 S--LRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAP-----YRDLSP 373
                |||       | |   |  : || | : :    ::   |  : :|      : :::
Db       1304 GEYLRSQGLRAMDPDRAVEELHITLDHGQTSVSVVDMDRRRFVELFTAARHRPLFDEIAG 1363

Qy        374 LGAEWAEKADGN--VRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEE 431
                 ||  :  :|    : :   :|      : :| ||  |      |  ||
Db       1364 ARAEARQSEEGPALAQRLAALLC----DGREREHLAHLIR--AEVAAVLG---------- 1407

Qy        432 HGS------DRWFNALGSRRFS----RNACVVVS----CGTAV----TVDALTDDGHYL 472
                ||      || |  ||    :    ||    |:      | |    |:  | |  |||
Db       1408 HGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAAVTGVREAATVVFDHPTITRLAD--HYL 1464


RESULT 15
GLTD_ECOLI
ID   GLTD_ECOLI     STANDARD;      PRT;   471 AA.
AC   P09832;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamate synthase [NADPH] small chain (EC 1.4.1.13) (Glutamate
DE   synthase beta subunit) (NADPH-GOGAT) (GLTS beta chain).
GN   GLTD OR ASPB OR B3213.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=K12;
RX   MEDLINE=88152492; PubMed=3326786;
RA   Oliver G., Gosset G., Sanchez-Pescador R., Lozoya E., Ku L.M.,
RA   Flores N., Becerril B., Valle F., Bolivar F.;
RT   "Determination of the nucleotide sequence for the glutamate synthase
RT   structural genes of Escherichia coli K-12.";
RL   Gene 60:1-11(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE OF 1-12.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=89098858; PubMed=2643092;
RA   Gosset G., Merino E., Recillas F., Oliver G., Becerril B., Bolivar F.;
RT   "Amino acid sequence analysis of the glutamate synthase enzyme from
RT   Escherichia coli K-12.";
RL   Protein Seq. Data Anal. 2:9-16(1989).
CC   -!- CATALYTIC ACTIVITY: 2 L-glutamate + NADP(+) = L-glutamine + 2-
CC       oxoglutarate + NADPH.
CC   -!- COFACTOR: BINDS A 4FE-4S CLUSTER.
CC   -!- PATHWAY: NITROGEN METABOLISM, GLUTAMATE BIOSYNTHESIS.
CC       THE CATALYZED REACTION BRINGS TOGETHER THE NITROGEN AND
CC       CARBON METABOLISM.
CC   -!- SUBUNIT: AGGREGATE OF 4 CATALYTICAL ACTIVE HETERODIMERS,
CC       CONSISTING OF A LARGE AND A SMALL SUBUNIT.
CC   -!- MISCELLANEOUS: GLUTAMINE BINDS TO THE LARGE SUBUNIT AND TRANSFERS
CC       THE AMIDO GROUP TO 2-OXO-GLUTAMATE THAT APPARENTLY BINDS TO THE
CC       SMALL SUBUNIT.
CC   -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC       'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC   -!- SIMILARITY: STRONG, TO E.COLI AEGA AND YGFT.
CC   ---------------------------------------------------------------------
```

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M18747; AAA23905.1; -.
DR   EMBL; U18997; AAA58015.1; -.
DR   EMBL; AE000400; AAC76245.1; -.
DR   PIR; B29617; B29617.
DR   SWISS-2DPAGE; P09832; COLI.
DR   ECO2DBASE; F050.4; 6TH EDITION.
DR   EcoGene; EG10404; gltD.
DR   InterPro; IPR000759; Adrndx_reductase.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00070; pyr_redox; 1.
DR   PRINTS; PR00419; ADXRDTASE.
DR   ProDom; PD000139; FAD_pyr_redox; 1.
DR   TIGRFAMs; TIGR01318; gltD_gamma_fam; 1.
KW   Oxidoreductase; Glutamate biosynthesis; Iron-sulfur; 4Fe-4S; NADP;
KW   Complete proteome.
FT   INIT_MET      0      0
FT   METAL        93     93       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL        97     97       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       103    103       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL       107    107       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   CONFLICT     37     50       GQAKAQADRCLSCG -> ARPKRRLTAACRAA (IN
FT                                REF. 1).
FT   CONFLICT    122    122       E -> K (IN REF. 1).
FT   CONFLICT    173    173       V -> C (IN REF. 1).
FT   CONFLICT    256    269       VYAALPFLIANTKQ -> CTQRCRSSSPTPNS (IN
FT                                REF. 1).
FT   CONFLICT    311    312       KH -> ND (IN REF. 1).
FT   CONFLICT    375    399       GRRRAEIVAGSEHIVPADAVIMAFG -> ASPRGDRCRFRT
FT                                YRTGRCGDHGVW (IN REF. 1).
SQ   SEQUENCE   471 AA;  51884 MW;  C4070F7509C561A8 CRC64;

  Query Match             3.5%;  Score 108;  DB 1;  Length 471;
  Best Local Similarity   22.4%;  Pred. No. 1.6;
  Matches  100;  Conservative   43;  Mismatches  128;  Indels  176;  Gaps   24;

Qy        139 FSFGWVFDRPQY----ELGSLSPV-----AAVACRRALSRLGLDVQIKWPNDLVVGR-DK 188
              |  ||   ||      : |    :     | :||   |:| |:         :|  |  :
Db        130 FEMGW---RPDMSGVKQTGKKVAIIGAGPAGLACADVLTRNGVKA-------VVFDRHPE 179

Qy        189 LGGILI---------ETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNAD 239
              :||:|         : | |  : :   |:|| | |  ||
Db        180 IGGLLTFGIPAFKLEKEVMTRRREI-FTGMGIEFKLNTEV------------------G 219

Qy        240 AAVLLETLLVELDAVLL-----QYARDGF-----------APFVAEYQAANRDHGKAVLL 283
                | |: || : ||| |     |  | |              ||:     ||         |
Db        220 RDVQLDDLLSDYDAVFLGVGTYQSMRGGLENEDADGVYAALPFL----IANTKQ-----L 270

Qy        284 LRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSG---------EISLRSDDRPVSVPKR 334
              :  |||  |  |                : |||: || |           |:|   : |:   |
Db        271 MGFGETRDEPFV-----------SMEGKRVVVLGGGDTAMDCVRTSVRQGAKHVTCAYR 318

Qy        335 RDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAV- 393
              || |    : |    :| |  |   |          :: ||| |     :|| :: |  :
Db        319 RDEEN---MPGSRREVKNAREEGVEFKF-------NVQPLGIE----VNGNGKVSGVKMV 364

Qy        394 ---CGE--FKKAQVQEQLARKIEWLPSSA--QALGIRNH-----YRHPEE---------- 431
                  ||   |  :  | :|     :|: |   | | | |      :|  |
Db        365 RTEMGEPDAKGRRRAEIVAGSEHIVPADAVIMAFGFRPHNMEWLAKHSVELDSQGRIIAP 424

Qy        432 HGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT 491
               |||  |     : |:                        || |: |  |:  ::|
Db        425 EGSDNAFQTSNPKIFA----------------------GGDIVRGSDLVVTAIA--- 456

Qy        492 ANLNRHAGKRYPFPTTTGNAVASGMMD 518
```

```
                                    |   | |:|:
Db        457 ----------------EGRKAADGIMN 467

Search completed: June 24, 2003, 21:51:37
Job time : 13.2965 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 21.5927 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-43
Perfect score: 3078
Sequence: 1 MTVLKLSHWRVLAELADGLP.....LVIYGLLNMIAAEGREYEHI 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3078	100.0	592	2	B81009	BirA protein/Bvg a
2	3024	98.2	592	2	H82031	probable biotin-[a
3	308.5	10.0	321	2	D91241	hypothetical prote
4	308.5	10.0	321	2	A86089	hypothetical prote
5	307.5	10.0	321	1	BVECBF	biotin-[acetyl-CoA
6	294.5	9.6	320	2	E82337	birA bifunctional
7	279.5	9.1	312	2	A83112	BirA bifunctional
8	275.5	9.0	320	2	AG0934	bifunctional prote
9	274.5	8.9	319	2	AG0457	biotin-[acetyl-CoA
10	254.5	8.3	267	2	I40327	baf protein - Bord
11	244	7.9	474	2	D75285	birA bifunctional
12	235.5	7.7	317	2	B82637	bifunctional bioti
13	234.5	7.6	248	2	H83111	hypothetical prote
14	223.5	7.3	274	2	E87000	biotin apo-protein
15	222.5	7.2	325	2	A69595	transcription repr
16	218.5	7.1	325	2	AH1684	transcription regu
17	218	7.1	323	2	E83860	biotin operon tran
18	217.5	7.1	325	2	AH1312	transcription regu
19	208.5	6.8	274	2	S74680	biotin-[acetyl-CoA
20	208	6.8	266	2	H70979	probable birA prot
21	203	6.6	233	2	A70351	biotin [acetyl-CoA
22	202.5	6.6	261	2	H69122	biotin acetyl-CoA
23	200	6.5	242	2	A82637	conserved hypothet
24	195.5	6.4	237	2	H75202	biotin operon repr
25	195	6.3	235	2	A71236	probable biotin-[a
26	194	6.3	311	2	B95222	BirA bifunctional
27	194	6.3	311	2	A99086	bioti-[acetyl-CoA-
28	192.5	6.3	240	2	B47751	biotin [acetyl-CoA
29	190.5	6.2	323	2	H89923	hypothetical prote
30	183.5	6.0	250	2	C87489	biotin protein lig
31	181	5.9	273	2	E97293	probable transcrip
32	179.5	5.8	231	2	C72401	biotin-(acetyl-CoA
33	177	5.8	265	2	T36391	hypothetical prote
34	175.5	5.7	256	2	AD2734	birA bifunctional
35	175.5	5.7	256	2	F97515	birA protein (AF20
36	174.5	5.7	268	2	D71657	biotin-protein lig
37	174.5	5.7	282	2	B84224	biotin acetyl-CoA
38	173	5.6	274	2	AH2462	biotin acetyl-CoA
39	172	5.6	261	2	B87489	transcription acti
40	171	5.6	323	2	D86847	biotin-[acetyl-CoA
41	169	5.5	355	2	G84651	biotin holocarboxy
42	164	5.3	270	2	E97796	biotin-[acetyl-CoA
43	162.5	5.3	310	2	G86490	probable biotin ho
44	155	5.0	250	2	G86862	biotin-[acetyl-CoA
45	150	4.9	274	2	H86937	conserved hypothet

ALIGNMENTS

RESULT 1
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strai
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, I
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g72
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

Query Match 100.0%; Score 3078; DB 2; Length 592;
Best Local Similarity 100.0%; Pred. No. 1.8e-225;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592


RESULT 2
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitidis (
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
  Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g737911
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

  Query Match             98.2%;  Score 3024;  DB 2;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 2.2e-221;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              ||||||||||||||||||||||||||||||||||||:||||:|||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 3
D91241
hypothetical protein ECs4900 [imported] - Escherichia coli (strain O157:H7, substrai
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: D91241
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, (
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D91241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38323.1; PID:g13364376; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4900
C;Superfamily: birA bifunctional protein

  Query Match             10.0%;  Score 308.5;  DB 2;  Length 321;
  Best Local Similarity   31.3%;  Pred. No. 9.6e-16;
  Matches  104;  Conservative   51;  Mismatches   144;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |   :|       ||: |    |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|| :  ||:  |    :      |:|  :|:  ||   |:    ||   |
Db         59 -SLPEPIQLLNAEQI--LGQLDGGSVTVLPVIDSTNQYLLD--RIGELKSGDA-CVAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQI 176
                ||||:|||    |  |  |  |  :: |   :| || |  :      | :|| | |::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 4
A86089
hypothetical protein birA [imported] - Escherichia coli (strain O157:H7, substrain E
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A86089
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86089
A;Status: preliminary
A;Molecule type: DNA
```

```
A;Residues: 1-321 <STO>
A;Cross-references: GB:AE005174; NID:g12518899; PIDN:AAG59173.1; GSPDB:GN00145; UWGP:Z55
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: birA
C;Superfamily: birA bifunctional protein

  Query Match            10.0%;  Score 308.5;  DB 2;  Length 321;
  Best Local Similarity   31.3%;  Pred. No. 9.6e-16;
  Matches  104;  Conservative   51;  Mismatches  144;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |    :|        ||: |   |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|| :  ||:  |    :     |:|  :|:  ||   |:    ||   |
Db         59 -SLPEPIQLLNAEQI--LGQLDGGSVTVLPVIDSTNQYLLD--RIGELKSGDA-CVAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQI 176
               ||||:||||    |  |  |  |  :: |   :| || |  :     | :|| | |::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 5
BVECBF
biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Escherichia coli (strain K-12)
N;Contains: biotin operon repressor; biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 31-Dec-1988 #text_change 03-Jun-2002
C;Accession: B24029; B45727; D65204; A24989
R;Howard, P.K.; Shaw, J.; Otsuka, A.J.
Gene 35, 321-331, 1985
A;Title: Nucleotide sequence of the birA gene encoding the biotin operon repressor and b
A;Reference number: A91525; MUID:86006282; PMID:3899863
A;Accession: B24029
A;Molecule type: DNA
A;Residues: 1-321 <HOW>
A;Cross-references: GB:M10123; NID:g145430; PIDN:AAA23520.1; PID:g145432
R;Buoncristiani, M.R.; Howard, P.K.; Otsuka, A.J.
Gene 44, 255-261, 1986
A;Title: DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (
A;Reference number: A24989; MUID:87055242; PMID:3536662
A;Contents: annotation; domain information
A;Note: this paper reports mutational studies done to understand DNA-binding and enzymat
R;Song, W.J.; Jackowski, S.
J. Bacteriol. 174, 6411-6417, 1992
A;Title: Cloning, sequencing, and expression of the pantothenate kinase (coaA) gene of E
A;Reference number: A45727; MUID:93015690; PMID:1328157
A;Accession: B45727
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 226-321 <SON>
A;Cross-references: GB:M90071
A;Note: residues 246-247 are correct to paper
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D65204
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-321 <BLAT>
A;Cross-references: GB:AE000471; GB:U00096; NID:g1790404; PIDN:AAC76951.1; PID:g179(
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: birA
A;Map position: 89 min
C;Function: <ENZ>
A;Description: EC 6.3.4.15 [validated; MUID:99400973]; biotin-[acetyl-CoA-carboxylas
otin-accepting proteins
C;Function: <REP>
A;Description: acts as a biotin operon repressor [validated; MUID:99400973]; the rea
orepressor
A;Note: the switch of birA from a biotin ligase to a repressor reflects competition
C;Superfamily: birA bifunctional protein
C;Keywords: ATP; biotin metabolism; DNA binding; ligase; repressor; transcription re
F;1-26/Domain: DNA binding #status predicted <DNA>
F;83-119,189-235/Domain: biotin activation #status predicted <BIO>

  Query Match            10.0%;  Score 307.5;  DB 1;  Length 321;
  Best Local Similarity   31.0%;  Pred. No. 1.1e-15;
  Matches  103;  Conservative   53;  Mismatches  143;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |    :|        ||: |   |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|| :  ||:  |    :     |:|  :|:  ||   |:    ||   |
Db         59 -SLPEPIQLLNAKQI--LGQLDGGSVAVLPVIDSTNQYLLD--RIGELKSGDA-CIAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQI 176
               ||||:||||    |  |  |  |  :: |   :| || |  :     | :|| | |::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 6
E82337
birA bifunctional protein VC0319 [imported] - Vibrio cholerae (strain N16961 serogrc
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: E82337
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, F
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Seller
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: E82337
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-320 <HEI>
A;Cross-references: GB:AE004120; GB:AE003852; NID:g9654727; PIDN:AAF93492.1; GSPDB:G
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0319
A;Map position: 1
C;Superfamily: birA bifunctional protein

  Query Match            9.6%;  Score 294.5;  DB 2;  Length 320;
```

```
  Best Local Similarity   28.9%;  Pred. No. 1.1e-14;
  Matches   98;  Conservative   45;  Mismatches  145;  Indels   51;  Gaps   11;

Qy         11 VLAELADGLPQHVSQ-------LARMADMKPQQLNGFWQQMPAHIRGLLR--------QH 55
              :| :|||   |  :       ::| |  |              ||:|:          |
Db         10 ILKQLADG-DFHSGEVLGAQLGISRAAISK------------HIQGIRDWGVDVFRVQG 55

Qy         56 DGYWRLVRPLAVFDAEGLRELGERSGFQTALKHE------CASSNDEILELARIAPDKAH 109
               || :| : : : |          :|  |: : :         |:|  :|:
Db         56 KGY-QLAQAMTLLD---------QSVIQSQVNNPVELHPIIGSTNQYLLDHVETL---VS 102

Qy        110 KTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSR 169
              |:|:   |:|||||:|| |    |  |  |  |  |         || |  ||   ||
Db        103 GTVCLAEYQASGRGRRGRHWVSPFGANLYLSIYWRLDAGMAAAMGLSLVVGVAIVEALEA 162

Qy        170 LGLD-VQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQS 227
              :|:| |::|||||      || |||:| : : |     |:|:|||  :     |
Db        163 MGVDGVKLKWPNDLYYQDKKLAGILVEMSGQAGAAAHLVIGMGINLAMRDNEGNIDQPWI 222

Qy        228 LFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDG 287
              : :   |   |  |:  ||  | ||   |  || :   :   |: | ||  |
Db        223 SLAEVTGQSRIDRNALAINLIAALDRTLRQYEISGMQNFVERWNRWDNFIGRPVKLLM-G 281

Qy        288 ETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDD 326
                  |  :|:|  | : ||| || :: : ||||||.:|
Db        282 ANEVRGIERGIDEHGGVLLETEEGLKSFIGGEISLRKND 320


RESULT 7
A83112
BirA bifunctional protein PA4280 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83112
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83112
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07668.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: birA; PA4280
C;Superfamily: birA bifunctional protein

  Query Match             9.1%;  Score 279.5;  DB 2;  Length 312;
  Best Local Similarity   33.1%;  Pred. No. 1.5e-13;
  Matches   90;  Conservative   38;  Mismatches  125;  Indels   19;  Gaps    8;

Qy         57 GYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTH 116
              || ||  ||:: |   :  |     :    :     |:| | :   |:|   :   : |
Db         52 GY-RLASPLSLLDPRKIDSL-----WPVEIMPTVDSTNSEAMR--RLAAGASTPFVIVAE 103

Qy        117 LQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-DVQ 175
               |: ||||:|| |    || | :|          ||  || |  :|  |||  ||: |
Db        104 QQTAGRGRRGRAWVSPYGENLYYSLAITVRGGARELEGLSLVVGLAVLRALESLGVQDTG 163

Qy        176 IKWPNDLVVGRDKLGGILIETVRTGGKTV----AVVGIGINFVLPKEVENAASVQSLFQT 231
              :||||||: |  |: |||:|     |        |:|||||  :    :      || :
Db        164 LKWPNDLLQGGKKIAGILLE---LSGDPADLCHVVIGIGINVNMKTATDIGQPWTSLREA 220

Qy        232 ASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                  | : ||  |  :|   | :::|:|||    |::: |   | | ::|    :
Db        221 LGEL--VDRSRLLSALNRQLADYLDRHSREGFAASREEWESCNLWQG-AQVILSTALSSI 277

Qy        292 EGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              || | ||| :| | ||    :|    ||:|||
Db        278 EGRVLGVDSRGALRLEVDGVEQGFSGGELSLR 309


RESULT 8
AG0934
bifunctional protein, biotin operon repressor and biotin-[acetyl-CoA carboxylase] sy
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG0934
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churc
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
A;Reference number: AB0502; PMID:11677608
A;Accession: AG0934
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-320 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD09497.1; PID:g16504614; GSPDB:GN00176
C;Genetics:
A;Gene: STY3741
C;Superfamily: birA bifunctional protein

  Query Match             9.0%;  Score 275.5;  DB 2;  Length 320;
  Best Local Similarity   29.9%;  Pred. No. 3e-13;
  Matches   98;  Conservative   50;  Mismatches  147;  Indels   33;  Gaps   13;

Qy         11 VLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GYW 59
              ::: ||||    ||      |    :|        ||: |    |           ||
Db         10 LISLLADGEFHSGEQLGERLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY- 58

Qy         60 RLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQS 119
               |  |: : ||: :    :      |:     |:|  :|:  ||   ::    ||   |
Db         59 SLPEPIQLLDADRIH--SQLDSGNVAVLPVIDSTNQYLLD--RIGELRSGDA-CVAEYQQ 113

Qy        120 KGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQIK 177
              ||||:||||    |  |  |  |  :: |   :| || |  :     |  || | |::|
Db        114 AGRGRRGRKWFSPFGANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRELGADKVRVK 172

Qy        178 WPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRG 236
              |||||  :   || |||:| | :||     |:| |||  : : ||     |
Db        173 WPNDLYLLDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGI 231

Qy        237 NADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVK 296
                |  :|   |: :| | |  : ::| :|::: ::  :    : | |:   : :| |  :
Db        232 TLDRNMLAAKLIYKLRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLIIGDKEIF-GISR 290

Qy        297 GVDGQGVLHLETAEGKQTVVSGEISLRS 324
              |:| |'| | ||     :  : |||||||
Db        291 GIDTQGALLLEQDGVIKPWMGGEISLRS 318


RESULT 9
AG0457
biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) [imported] - Yersinia pestis (s
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C;Accession: AG0457
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, I
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barre
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AG0457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-319 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93227.1; PID:g15981675; GSPDB:GN00175
C;Genetics:
A;Gene: birA
```

```
C;Superfamily: birA bifunctional protein
C;Keywords: ligase

  Query Match             8.9%;  Score 274.5;  DB 2;  Length 319;
  Best Local Similarity   29.0%;  Pred. No. 3.6e-13;
  Matches   96;  Conservative   46;  Mismatches  150;  Indels   39;  Gaps   12;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFD 69
              |::: |:||      ||     |   :|   |     ||          |     | ||
Db          9 RLVSILSDGFFHSGEQLGETLGMSRAAINKHMQT----IRD--------W----GLDVFT 52

Qy         70 AEG--------LRELGER-------SGFQTALKHECASSNDEILELARIAPDKAHKTICV 114
                |        :: | |:        :|  | |     |:|  :|:  ||   |:    ||
Db         53 VPGKGYSLPTSIQLLDEQKILSYLPTGHVTVLP-VVDSTNQYLLD--RITELKSGDA-CV 108

Qy        115 THLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLD- 173
                |  ||||:||:|    |  |  |  |  ::        || |  :     |  || :
Db        109 AEYQQAGRGRRGRQWVSPFGANLYLSMYWRLEQGPAAAMGLSLVVGIVMAEVLQNLGAEK 168

Qy        174 VQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTA 232
              |::|||||| :   || |||:| | :||     |:| |||  :     || : |
Db        169 VRVKWPNDLYLNDKKLAGILVELTGKTGDAAQLVIGAGINLTMRGSETNAIN-QDWINLQ 227

Qy        233 SRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFE 292
                  |   |   :| ||   ::::  :| : |:: :|  :    : | |:   : :|
Db        228 DAGVTIDRNKLTAEILSELRLAVVKFENEGLSAFLSRWQKMDNYLNRPVKLIIGNQEIF- 286

Qy        293 GTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              |  :|:| || | ||     :  : ||||||
Db        287 GIARGIDQQGALLLEQDGNIKPYIGGEISLR 317


RESULT 10
I40327
baf protein - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40327; S70669
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
J. Bacteriol. 177, 3801-3807, 1995
A;Title: Identification of a Bordetella pertussis regulatory factor required for transcr
A;Reference number: I40327; MUID:95325323; PMID:7601846
A;Accession: I40327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229
A;Allen, A.; Maskell, D.
Mol. Microbiol. 19, 37-52, 1996
A;Title: The identification, cloning and mutagenesis of a genetic locus required for lip
A;Reference number: S70669; MUID:96419162; PMID:8821935
A;Accession: S70669
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 239-267 <ALL>
A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: baf

  Query Match             8.3%;  Score 254.5;  DB 2;  Length 267;
  Best Local Similarity   31.6%;  Pred. No. 9.3e-12;
  Matches   84;  Conservative   44;  Mismatches  113;  Indels   25;  Gaps    9;

Qy        341 LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 395
              :|:| ||||||  | : :   |    ||      ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        396 EFKKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 449
                :   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        450 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTG 509
                :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        510 NAVASGMMDAVCGSVM--MMHGRLKEKTGAGKPVDVIITGGG-------AAKVAEALPPA 560
                |:|||:  |  |:::   : || :     |:  :: : |||       | ::      |
Db        183 QAIASGIAAAQAGAIVRQWLAGRQR----YGQAPEIYVAGGGWPEVRQEAERLLAVTGAA 238

Qy        561 FLAENTVRVADNLVIYGLLNMIAAEG 586
              | |       |: |: ||   :||:|
Db        239 FGATPQPTYLDSPVLDGLA-ALAAQG 263


RESULT 11
D75285
birA bifunctional protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: D75285
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75285
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-474 <WHI>
A;Cross-references: GB:AE002065; GB:AE000513; NID:g6460149; PIDN:AAF11888.1; PID:g64
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2341
A;Map position: 1

  Query Match             7.9%;  Score 244;  DB 2;  Length 474;
  Best Local Similarity   32.5%;  Pred. No. 1.3e-10;
  Matches  111;  Conservative   42;  Mismatches  121;  Indels   68;  Gaps   20;

Qy          9 WRVLAELADGL---PQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPL 65
              ||: | |    |   ||    | |  |:    :|   :::          | ||      |:
Db        167 WRMPARLLPLLTDRPQSGDDLGRALDLGRVTVNTLARRL---------QEDGV-----PV 212

Qy         66 AVFDAEGLRELGE--------RSGFQTALKHE--CASSNDEILELARIAPDK-----AHK 110
                :  |    | |          |  |  ||:::    |: ||:    |   |       |
Db        213 LISRAGYALEPGTPAPGLVPLRGQFGRALRYQGTVTSTQDEV----RFWADSERGPAPHG 268

Qy        111 TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSL--SPVAA-VACRRAL 167
                : |   |: ||||:|| |    |  | ||   |  |    |   |:|| ||   |
Db        269 AVVVAERQTAGRGRRGRVWDTTHG-TLAFS---VLLREALSLADLVRMPLAAGVALHAAC 324

Qy        168 SRLGLDVQIKWPNDLVV--GRDKLGGILIETVRTGGKT-VAVVGIGINFVLPKEVENAAS 224
              : ||:   :||||||:   || || |||:|    | :   ||:|||||  :    | ||
Db        325 TALGVPCGLKWPNDLLTPDGR-KLAGILLEADLRGEEARRAVLGIGIN--VGTAPEGAAH 381

Qy        225 VQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLL 284
              :       :|     |||| | :| :|:  | | |       :| ::| :   |: | :
Db        382 LHESVPDLTR-----AAVLGE-VLFQLEHWLAQPA----PAIIATWKATSLTLGRPVRV- 430

Qy        285 RDGETVFEGTVKGV----DGQGVLHLETAEGKQTVVSGEISL 322
                  |   |||:||    | || | :||| | :|| :|:: |
Db        431 ----TTARGTVEGVARDLDEQGNLLVETAAGLETVSAGDVQL 468


RESULT 12
B82637
bifunctional biotin operon/biotin acetyl-CoA-carboxylase synthetase transcription re
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: B82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
```

```
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-317 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84604.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silvei
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
C;Contents: annotation
C;Genetics:
A;Gene: XF1796
C;Superfamily: birA bifunctional protein

  Query Match             7.7%;  Score 235.5;  DB 2;  Length 317;
  Best Local Similarity   28.6%;  Pred. No. 3.3e-10;
  Matches   83;  Conservative   46;  Mismatches  120;  Indels   41;  Gaps   11;

Qy         53 RQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECA----------SSNDEILE 99
              |   ||  |  ||:: :|| :|           | |  |             |:|  :|
Db         48 RAEQGY-ALAAPLSLLNAERIR---------AELSHSIAAELRVLDVAWTLESTNTSLLA 97

Qy        100 LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVA 159
                :|      : :   |: |||::|| |   |   :  |   ||      :|  ||
Db         98 RPALADG---VEVLLAERQTGGRGQRGRIWISPLAANIYLSLSRVFSSGAAQLMGLSVAV 154

Qy        160 AVACRRALSRLGLD-VQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLP-- 216
               ||   ||   |   |::|||||||    ||||:|:|     |:  ||:|:|:|  :|
Db        155 GVAIAEALIAAGFSHVRLKWPNDLVDNDRKLGGVLVE---RQGEARAVIGLGLNVRMPTN 211

Qy        217 --KEVENA-ASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAA 273
                 ::::     :::|  :   | ||  |:||  ||  |    |    :| |||:  |
Db        212 FAEQIDQPWVDLETLAGSTVER-NALVALLLSALLPALHLFQL----EGLAPFLPRYAKL 266

Qy        274 NRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              :   | ||  : :    : |  :|:   | | :|     :   :|::|:|
Db        267 DCLVGCAV-WIEEANLKYVGIARGLAADGALLVEINGVLRPFHTGQVSVR 315


RESULT 13
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

  Query Match             7.6%;  Score 234.5;  DB 2;  Length 248;
  Best Local Similarity   27.6%;  Pred. No. 2.8e-10;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Qy        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453
              :  : :||  :  | ||        |  |:|| |   :  | |||   : :   :: ||:|
Db         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Qy        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |     |
Db        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Qy        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561
                | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

Qy        562 LAENTVRVADNLVIYGL 578
                  |:  :||  ||
Db        231 ------RIMPDLVFVGL 241


RESULT 14
E87000
biotin apo-protein ligase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 17-May-2002
C;Accession: E87000
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: E87000
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092858; PIDN:CAC30241.1; GSPDB:GN00147
C;Genetics:
A;Gene: birA
C;Superfamily: Bacillus subtilis transcription repressor of biotin operon birA

  Query Match             7.3%;  Score 223.5;  DB 2;  Length 274;
  Best Local Similarity   34.9%;  Pred. No. 2.2e-09;
  Matches   87;  Conservative   32;  Mismatches  101;  Indels   29;  Gaps   12;

Qy         89 ECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFG-WVFDR 147
              :  |:| ::|  |    |      :   | |: |||| || ||       :  | |  : |
Db         34 QTGSTNADLLARAASGIDIDGAVLIAEH-QTAGRGRHGRGWSASPRTQITMSIGVSIADV 92

Qy        148 PQYELGSLSPVAAVACRRALSRL----GLDVQIKWPNDLVVG----RDKLGGILIETVRT 199
              |      ||    ||    :: |    |:|  :|||||:: |    : || ||| | ||
Db         93 PVATWAWLSLATGVAVLDTVAPLIDVTGVDAGLKWPNDVLAGPPGSKGKLAGILAEVVR- 151

Qy        200 GGKTVAVVGIGINFV-LPKEVE--NAASVQSLFQTASRRGNADAAVLLETLLVELDAVLL 256
                | |||:|:|    ||||:   | |:  |   |      |   |:  || |||| :
Db        152 ---PVIVVGVGLNVTQAPKEVDGPGATSLLDLGVPA-----PDREQLVCWLLRELDARVG 203

Qy        257 QYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK---Q 313
              |: |     :|:|:| :   |  |     |     |||:|:| || | |:| ||    |
Db        204 QW-RSADPQLMADYRARSPIIGSRVRAELPGGKDVVGTVRGIDDQGRLCLKT-EGPTKGQ 261

Qy        314 TVV--SGEI 320
              |||  :|::
Db        262 TVVISAGDV 270


RESULT 15
A69595
transcription repressor of biotin operon birA - Bacillus subtilis
```

C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: A69595
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69595
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-325 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14160.1; PID:g2634662
A;Experimental source: strain 168
C;Genetics:
A;Gene: birA
C;Superfamily: Bacillus subtilis transcription repressor of biotin operon birA

Query Match 7.2%; Score 222.5; DB 2; Length 325;
Best Local Similarity 27.7%; Pred. No. 3.3e-09;
Matches 85; Conservative 50; Mismatches 119; Indels 53; Gaps 13;

```
Qy         47 HIRGLLRQ--------HDGYWRLVRPLAVFDAE---GLRE--LGERSGFQTALK------ 87
              ||  | ::          ||  : :|  : ::|   ||:   :|:   :   |
Db         40 HIEELRKEGYEVEAVRRKGYRLIKKPGKLSESEIRFGLKTEVMGQHLIYHDVLSSTQKTA 99

Qy         88 HECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDR 147
              || |::|         ||:    |: |   |: ||||  | |  : |  :  |     |
Db        100 HELANNN---------APE---GTLVVADKQTAGRGRMSRVWHSQEGNGVWMSLILRPDI 147

Qy        148 PQYELGSLSPVAAVACRRALSR-LGLDVQIKWPNDLVVGRDKLGGILIE------TVRTG 200
              |  :   |: :||||  : :    |:   ||||||:::   |  ||| |       ||:
Db        148 PLQKTPQLTLLAAVAVVQGIEEAAGIQTDIKWPNDILINGKKTVGILTEMQAEEDRVRS- 206

Qy        201 GKTVAVVGIGINF-----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVL 255
                 ::|||||        | |::: |:  || | |  :   | | ::: :|:  :
Db        207 ----VIIGIGINVNQQPNDFPDELKDIAT--SLSQAAGEK--IDRAGVIQHILLCFEKRY 258

Qy        256 LQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTV 315
              |   || |    :::     |   :  |     | |   |:| :||| ||| || : :
Db        259 RDYMTHGFTPIKLLWESYALGIG-TNMRARTLNGTFYGKALGIDDEGVLLLETNEGIKKI 317

Qy        316 VSGEISL 322
              | :| |
Db        318 YSADIEL 324
```

Search completed: June 24, 2003, 22:03:23
Job time : 22.5927 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 60.7795 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-43
Perfect score: 3078
Sequence: 1 MTVLKLSHWRVLAELADGLP.....LVIYGLLNMIAAEGREYEHI 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3078	100.0	592	20	AAY38615	Neisseria meningit
2	3078	100.0	592	21	AAY74912	Neisseria meningit
3	3078	100.0	592	23	AAU91169	Pantothenate kinas
4	3024	98.2	592	20	AAY38616	Neisseria meningit
5	3024	98.2	592	21	AAY74913	Neisseria meningit
6	3024	98.2	592	23	AAU91166	Pantothenate kinas
7	2885	93.7	592	20	AAY38618	Neisseria gonorrho
8	2885	93.7	592	21	AAY74911	Neisseria gonorrhe
9	2300	74.7	455	21	AAY74910	Neisseria meningit
10	2197	71.4	460	23	AAU91167	Pantothenate kinas
11	2171	70.5	455	20	AAY38617	Neisseria gonorrho
12	2171	70.5	455	21	AAY74908	Neisseria gonorrhe
13	1891.5	61.5	389	21	AAY74909	Neisseria meningit
14	973.5	31.6	189	20	AAY38614	Neisseria meningit
15	307.5	10.0	322	20	AAY31601	Escherichia coli b
16	272	8.8	363	22	ABG28654	Novel human diagno
17	259	8.4	302	22	AAB88546	Haemophilus influe
18	259	8.4	302	23	AAU91477	Haemophilus influe
19	254.5	8.3	267	23	AAU91162	Bordella pertussis
20	245	8.0	249	23	AAU91178	Pantothenate kinas
21	234.5	7.6	248	23	AAU91164	Pantothenate kinas
22	217.5	7.1	325	23	ABB48003	Listeria monocytog
23	214	7.0	291	22	AAU04901	Micromonospora eve
24	205.5	6.7	282	22	AAU64093	Propionibacterium
25	200	6.5	242	23	AAU91180	Pantothenate kinas
26	195.5	6.4	237	22	AAB96079	Putative P. abyssi
27	194	6.3	311	21	AAY70846	Streptococcus pneu
28	192	6.2	294	22	ABB09064	Thermus caldophilu
29	191	6.2	249	23	AAU91182	Pantothenate kinas
30	190.5	6.2	323	21	AAB07758	Amino acid sequenc
31	184.5	6.0	269	22	AAB76593	Corynebacterium gl
32	184.5	6.0	288	22	AAG90532	C glutamicum prote
33	177	5.8	265	23	AAU91151	Streptomyces coeli
34	175.5	5.7	323	22	ABB09053	Thermus caldophilu
35	172	5.6	260	23	AAU91173	Pantothenate kinas
36	171	5.6	323	23	ABB55129	Lactococcus lactis
37	169.5	5.5	262	23	AAU91170	Pantothenate kinas
38	169	5.5	254	21	AAG32612	Arabidopsis thalia
39	169	5.5	266	21	AAG06098	Arabidopsis thalia
40	169	5.5	266	21	AAG49376	Arabidopsis thalia
41	169	5.5	317	21	AAG32611	Arabidopsis thalia
42	169	5.5	329	21	AAG06097	Arabidopsis thalia
43	169	5.5	329	21	AAG49375	Arabidopsis thalia
44	169	5.5	355	21	AAG32610	Arabidopsis thalia
45	169	5.5	367	21	AAG06096	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY38615
ID AAY38615 standard; Protein; 592 AA.
XX
AC AAY38615;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX

```
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             100.0%;  Score 3078;  DB 20;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-282;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592


RESULT 2
AAY74912
ID   AAY74912 standard; Protein; 592 AA.
XX
AC   AAY74912;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53674.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 714; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             100.0%;  Score 3078;  DB 21;  Length 592;
  Best Local Similarity   100.0%;  Pred. No. 2.4e-282;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
```

```
Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
```

RESULT 3
AAU91169
ID AAU91169 standard; Protein; 592 AA.
XX
AC AAU91169;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #7.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria meningitidis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54190.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 96-98; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX

SQ Sequence 592 AA;

Query Match 100.0%; Score 3078; DB 23; Length 592;
Best Local Similarity 100.0%; Pred. No. 2.4e-282;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
```

RESULT 4
AAY38616
ID AAY38616 standard; Protein; 592 AA.
XX
AC AAY38616;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis strain A antigen encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.

```
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             98.2%;  Score 3024;  DB 20;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 3.1e-277;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

          301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 5
AAY74913
ID   AAY74913 standard; Protein; 592 AA.
XX
AC   AAY74913;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1300.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53675.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 715; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             98.2%;  Score 3024;  DB 21;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 3.1e-277;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 6
AAU91166
ID   AAU91166 standard; Protein; 592 AA.
XX
AC   AAU91166;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #4.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria meningitidis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 85-86; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   592 AA;

  Query Match             98.2%;  Score 3024;  DB 23;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 3.1e-277;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 7
AAY38618
ID   AAY38618 standard; Protein; 592 AA.
XX
AC   AAY38618;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigenic protein encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
```

```
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12075.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             93.7%;  Score 2885;  DB 20;  Length 592;
  Best Local Similarity   93.9%;  Pred. No. 4.6e-264;
  Matches  555;  Conservative   14;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              |||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              |||||||||||||||||||||| |||||||||||||||:|||||  |||: ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| || ||| ||| :|||||: ||: ||||||||||||||||| |||||||||
Db        241 AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              :|||||||||:|||||||||| |:| |||||| ||||||||:||||||||||||||||
Db        301 RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||||||||||||||||| |||||:|||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||| |||||||||||||||||||||||||:|||||||||| |||||
Db        481 HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 8
AAY74911
ID   AAY74911 standard; Protein; 592 AA.
XX
AC   AAY74911;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1296.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53673.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 713; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             93.7%;  Score 2885;  DB 21;  Length 592;
  Best Local Similarity   93.9%;  Pred. No. 4.6e-264;
  Matches  555;  Conservative   14;  Mismatches   22;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| |||||||||||||||||||| |||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
```

```
Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              |||||||||||||||||||||| ||||||||||||||:||||||  |||: ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| || ||| ||| :|||||: ||: |||||||||||||||| ||||||||||
Db        241 AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              :|||||||||||:|||||||||| |:| |||||| |||||||:|||||||||||||||||
Db        301 RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||||||||||||||| |||||:|||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              |||||||||||||||| ||||||||||||||||||||||||:||||||||-|| |||||
Db        481 HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 9
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             74.7%;  Score 2300;  DB 21;  Length 455;
  Best Local Similarity   98.2%;  Pred. No. 7.9e-209;
  Matches  446;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 197
              |||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db          1 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQ 257
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db         61 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 120

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 180

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        378 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 437
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        438 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 360

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 420

Qy        558 PPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||:||||:||||| | ||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 454


RESULT 10
AAU91167
ID   AAU91167 standard; Protein; 460 AA.
XX
AC   AAU91167;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #5.
```

```
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54188.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 8; Page 92-94; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   460 AA;

  Query Match            71.4%;  Score 2197;  DB 23;  Length 460;
  Best Local Similarity  92.6%;  Pred. No. 4.6e-199;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192
              :|||||||||| |||||||||||||||:||||||  |||: ||||||||||||||||||
Db          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60

Qy        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120

Qy        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||||:
Db        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180

Qy        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372
              |||||||||| |:| |||||| ||||||||:|||||||||||||||||||||||||||||
Db        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240

Qy        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432
              |||||||||||||||||||||| |||||:|||||||||||||||||||||||||||||||
Db        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300

Qy        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360

Qy        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552
              |||| |||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420

Qy        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||||:||||:||||| | ||
Db        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459


RESULT 11
AAY38617
ID   AAY38617 standard; Protein; 455 AA.
XX
AC   AAY38617;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   455 AA;

  Query Match            70.5%;  Score 2171;  DB 20;  Length 455;
  Best Local Similarity  92.7%;  Pred. No. 1.3e-196;
  Matches  421;  Conservative   13;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 197
              |||||| |||||||||||||:|||||  |||: |||||||||||||||||||||||||
Db          1 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQ 257
              | |||||||||||||||||||||||||||||||||||||||||||||||||| || ||| |
Db         61 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 120

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
              || :|||||: ||: |||||||||||||||| |||||||||:||||||||||:|||||
Db        121 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 180

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
```

```
                ||||||  |:|  ||||||  ||||||||:|||||||||||||||||||||||||||||||||||
Db        181 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        378 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 437
                ||||||||||||||||||  |||||:|||||||||||||||||||||||||||||||||||
Db        241 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        438 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 360

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
                ||||||||||||||||||||||||:||||||||| ||||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 420

Qy        558 PPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
                ||||||||||||||||:||||:|||||  |  ||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 454


RESULT 12
AAY74908
ID   AAY74908 standard; Protein; 455 AA.
XX
AC   AAY74908;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1290.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53670.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 709; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             70.5%;  Score 2171;  DB 21;  Length 455;
  Best Local Similarity   92.7%;  Pred. No. 1.3e-196;
  Matches  421;  Conservative   13;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 197
                ||||||  ||||||||||||||||:||||||  |||: ||||||||||||||||||||||
Db          1 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQ 257
                | ||||||||||||||||||||||||||||||||||||||||||||||||| || ||| |
Db         61 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 120

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
                || :|||||: ||: ||||||||||||||||| |||||||||:||||||||||:|||||
Db        121 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 180

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
                ||||||  |:|  ||||||  ||||||||:|||||||||||||||||||||||||||||||
Db        181 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        378 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 437
                ||||||||||||||||||  |||||:|||||||||||||||||||||||||||||||||||
Db        241 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        438 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 360

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
                ||||||||||||||||||||||||:||||||||| ||||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 420

Qy        558 PPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
                ||||||||||||||||:||||:|||||  |  ||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 454


RESULT 13
AAY74909
ID   AAY74909 standard; Protein; 389 AA.
XX
AC   AAY74909;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
```

```
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53671.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 710; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   389 AA;

  Query Match             61.5%;  Score 1891.5;  DB 21;  Length 389;
  Best Local Similarity   82.9%;  Pred. No. 3.1e-170;
  Matches  377;  Conservative    0;  Mismatches   11;  Indels   67;  Gaps    2;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQ 257
              ||||||||||||||||||| ||||||||||||||||||||||||
Db         61 RTGGKTVAVVGIGINFVLPXEVENAASVQSLFQTASRRGNADAAVLLXX---------- 109

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317

Db        110 ----------------------------------------------------XXXXX 114

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
              ||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        115 XEISLRSDXRPVSVXKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 174

Qy        378 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL-GIRNHYRHPEEHGSDR 436
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        175 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALFGIRNHYRHPEEHGSDR 234

Qy        437 WFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 496
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 WFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 294

Qy        497 HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEA 354

Qy        557 LPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              ||||||||||||||||||||||||||||||||||
Db        355 LPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 389

RESULT 14
```

```
AAY38614
ID   AAY38614 standard; Protein; 189 AA.
XX
AC   AAY38614;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12072.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 171; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   189 AA;

  Query Match             31.6%;  Score 973.5;  DB 20;  Length 189;
  Best Local Similarity   97.9%;  Pred. No. 9e-84;
  Matches  185;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy        319 EISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEW 378
              ||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 EISLRSDXRPVSVXKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEW 60

Qy        379 AEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWF 438
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 AEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAXGIRNHYRHPEEHGSDRWF 120

Qy        439 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLG-GTIMPGFHLMKESLAVRTANLNRH 497
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        121 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGXGTIMPGFHLMKESLAVRTANLNRH 180

Qy        498 AGKRYPFPT 506
              |||||||||
Db        181 AGKRYPFPT 189

RESULT 15
```

AAY31601
ID AAY31601 standard; Protein; 322 AA.
XX
AC AAY31601;
XX
DT 09-NOV-1999 (first entry)
XX
DE Escherichia coli biotin synthesis repressor enzyme BirA.
XX
KW BirA; DNA primer; PCR; polymerase chain reaction;
KW biotin synthesis repressor; biotin.
XX
OS Escherichia coli.
XX
PN WO9937785-A1.
XX
PD 29-JUL-1999.
XX
F 25-JAN-1999; 99WO-GB00245.
X
PR 26-JAN-1998; 98GB-0001621.
XX
PA (ISIS-) ISIS INNOVATION LTD.
XX
PI Jakobsen BK, O'Callaghan CA;
XX
DR WPI; 1999-469137/39.
DR N-PSDB; AAZ08957.
XX
PT Production of a bacterial repressor (BirA enzyme), useful for the
PT specific labeling of proteins with biotin
XX
PS Claim 6; Fig 1; 14pp; English.
XX
CC The present sequence is an Escherichia coli BirA enzyme which is a
CC biotin biosynthesis repressor. DNA primers were used to amplify the BirA
CC gene using DNA derived from E. coli, for subsequent cloning of the gene
CC into plasmid pGEX-2T, containing a thrombin cleavage recognition sequence
CC between a sequence encoding a glutathione-S-transferase tag and the
CC cloning site into which the BirA cloning sequence was inserted. The
CC resulting plasmid, encoding BirA as a fusion protein, was designated
CC COC 053. The plasmid was then used to transform E. coli
CC BL21(DE3) pLysS(4) for recombinant BirA enzyme production. The method
CC is used for specific transfer of a biotin group to a protein containing
CC a specific target sequence, in order to specifically label an engineered
CC protein with biotin. The method also produces higher quality amounts
CC of the BirA enzyme compared to existing methods.
XX
SQ Sequence 322 AA;

Query Match 10.0%; Score 307.5; DB 20; Length 322;
Best Local Similarity 31.0%; Pred. No. 2.5e-20;
Matches 103; Conservative 53; Mismatches 143; Indels 33; Gaps 13;

```
Qy          10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
               :::| ||:|      ||     |    :|        ||: |    |           ||
Db          10 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 59

Qy          59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                 |  |: : :|: :  ||:  |   |:     |:|  :|:  ||   |:    |:   |
Db          60 -SLPEPIQLLNAKQI--LGQLDGGSVAVLPVIDSTNQYLLD--RIGELKSGDA-CIAEYQ 113

Qy         119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGLD-VQI 176
                ||||:||||    |  |  |  |  :: |   :| || |  :     | :|| | |::
Db         114 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 172

Qy         177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
               |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db         173 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 231

Qy         236 GNADAAVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
                | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db         232 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 290

Qy         296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
               :|:| || | ||     :  : |||||| ::
Db         291 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 322
```

Search completed: June 24, 2003, 21:46:30
Job time : 61.7795 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 26.9909 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-22
Perfect score: 3080
Sequence: 1 MTVLKPSHWRVLAELADGLP.....LVIHGLLNLIAAEGGESEHT 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3080	100.0	592	9	US-09-813-453A-22	Sequence 22, Appl
2	3024	98.2	592	9	US-09-813-453A-43	Sequence 43, Appl
3	2223	72.2	460	9	US-09-813-453A-39	Sequence 39, Appl
4	256.5	8.3	267	9	US-09-813-453A-15	Sequence 15, Appl
5	256	8.3	302	9	US-10-260-877-110	Sequence 110, App
6	243	7.9	249	9	US-09-813-453A-61	Sequence 61, Appl
7	231.5	7.5	248	9	US-09-813-453A-20	Sequence 20, Appl
8	198.5	6.4	242	9	US-09-813-453A-65	Sequence 65, Appl
9	187	6.1	249	9	US-09-813-453A-70	Sequence 70, Appl
10	183.5	6.0	288	9	US-09-738-626-4286	Sequence 4286, Ap
11	180.5	5.9	265	9	US-09-813-453A-4	Sequence 4, Appli
12	179	5.8	311	10	US-09-859-270-2	Sequence 2, Appli
13	175	5.7	260	9	US-09-813-453A-51	Sequence 51, Appl
14	168	5.5	250	9	US-09-813-453A-3	Sequence 3, Appli
15	168	5.5	262	9	US-09-813-453A-45	Sequence 45, Appl
16	159.5	5.2	257	9	US-09-813-453A-53	Sequence 53, Appl
17	156	5.1	244	9	US-09-813-453A-41	Sequence 41, Appl
18	152	4.9	219	9	US-09-813-453A-57	Sequence 57, Appl
19	152	4.9	272	9	US-09-712-363-276	Sequence 276, App
20	152	4.9	272	9	US-09-813-453A-5	Sequence 5, Appli
21	150.5	4.9	246	9	US-09-813-453A-9	Sequence 9, Appli
22	150	4.9	258	9	US-09-813-453A-2	Sequence 2, Appli
23	147.5	4.8	273	9	US-09-813-453A-10	Sequence 10, Appl
24	146	4.7	258	9	US-09-813-453A-49	Sequence 49, Appl
25	141	4.6	255	9	US-09-813-453A-7	Sequence 7, Appli
26	141	4.6	257	9	US-09-813-453A-13	Sequence 13, Appl
27	140.5	4.6	256	9	US-09-813-453A-55	Sequence 55, Appl
28	140	4.5	258	9	US-09-813-453A-6	Sequence 6, Appli
29	137	4.4	254	9	US-09-813-453A-47	Sequence 47, Appl
30	133	4.3	233	9	US-09-813-453A-17	Sequence 17, Appl
31	132.5	4.3	212	9	US-09-813-453A-59	Sequence 59, Appl
32	116.5	3.8	223	9	US-09-895-913A-74	Sequence 74, Appl
33	116.5	3.8	223	9	US-09-813-453A-14	Sequence 14, Appl
34	116.5	3.8	223	9	US-09-813-453A-67	Sequence 67, Appl
35	115	3.7	3816	9	US-09-808-880-3	Sequence 3, Appli
36	112.5	3.7	241	9	US-09-813-453A-63	Sequence 63, Appl
37	112	3.6	472	10	US-09-815-242-13727	Sequence 13727, A
38	110	3.6	340	10	US-09-815-242-11743	Sequence 11743, A
39	108.5	3.5	286	9	US-09-769-787-100	Sequence 100, App
40	108	3.5	229	9	US-09-813-453A-12	Sequence 12, Appl
41	106.5	3.5	335	10	US-09-815-242-5138	Sequence 5138, Ap
42	105	3.4	1222	10	US-09-137-531-15	Sequence 15, Appl
43	105	3.4	1252	10	US-09-137-531-9	Sequence 9, Appli
44	104.5	3.4	4613	9	US-09-860-846-31	Sequence 31, Appl
45	104.5	3.4	4613	9	US-09-988-384B-31	Sequence 31, Appl

ALIGNMENTS

RESULT 1
US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-813-453A-22

Query Match 100.0%; Score 3080; DB 9; Length 592;
Best Local Similarity 100.0%; Pred. No. 3.7e-251;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592


RESULT 2
US-09-813-453A-43
; Sequence 43, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-43

  Query Match           98.2%;  Score 3024;  DB 9;  Length 592;
  Best Local Similarity  98.5%;  Pred. No. 2e-246;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              ||||||||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 3
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39

  Query Match           72.2%;  Score 2223;  DB 9;  Length 460;
  Best Local Similarity  93.7%;  Pred. No. 4.5e-179;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192
              :|||||||||| ||||||||||||||:|||||  |||:|||||||||||||||||||||
Db          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60

Qy        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252
              ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120

Qy        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||:
Db        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180

Qy        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372
              |||||||||| |:| |||||| ||||||||:|||||||||||||||||||||||||||||
Db        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240

Qy        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432
              ||||||||| |||||||||||||  ||||||:|||||||||||||||||||||||||||||
Db        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300

Qy        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360

Qy        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552
              |||| |||||||||||||||||||||||||:|||||||||| ||||||||||||||||
Db        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420

Qy        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||||||||
Db        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459


RESULT 4
US-09-813-453A-15
; Sequence 15, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
   TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
   TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Bordetella pertussis
US-09-813-453A-15

  Query Match             8.3%;  Score 256.5;  DB 9;  Length 267;
  Best Local Similarity   32.0%;  Pred. No. 1.2e-13;
  Matches   85;  Conservative   43;  Mismatches  113;  Indels   25;  Gaps    9;

Qy        341 LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKVDGNVRIVGCAVCG 395
              :|:| ||||||  | : :   |    ||       ||  ||   |        | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        396 EFKKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 449
              :   :   |       | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        450 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTG 509
              :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        510 NAVASGMMDAVCGSVM--MMHGRLKEKTGAGKPVDVIITGGG-------AAKVAEALPPA 560
              |:|||:  |  |:::   : || :     |:  :: : |||       | ::      |
Db        183 QAIASGIAAAQAGAIVRQWLAGRQR----YGQAPEIYVAGGGWPEVRQEAERLLAVTGAA 238

Qy        561 FLAENTVRVADNLVIHGLLNLIAAEG 586
              | |       |: |: ||  | ||:|
Db        239 FGATPQPTYLDSPVLDGLAAL-AAQG 263


RESULT 5
US-10-260-877-110
; Sequence 110, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Chovan, Linda E.
;  APPLICANT:  Hessler, Paul E.
;  APPLICANT:  Reich, Karl A.
;  TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
;  TITLE OF INVENTION:  SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  'ESSENTIAL GENES'
;  FILE REFERENCE: 6565.US.P1
;  CURRENT APPLICATION NUMBER: US/10/260,877
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: US/09/649,145
;  PRIOR FILING DATE: 2000-08-25
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 110
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: H. influenzae
US-10-260-877-110

  Query Match             8.3%;  Score 256;  DB 9;  Length 302;
  Best Local Similarity   29.1%;  Pred. No. 1.5e-13;
  Matches   91;  Conservative   56;  Mismatches  132;  Indels   34;  Gaps   13;

Qy         26 LARMADMKPQ---QLNGFWQQMPAHIR-----GLLRQHDGY-WRLVRPLAVFDAEGLREL 76
              |  :|| :|:   :|  | : :   |:     ||    ||  :|||  | : : : :
Db          7 LTYLADCQPKVRSELEKFSKNLEEDIQQLREIGLDILVDGQDYRLVPMLPLLNPQQI--- 63

Qy         77 GERSGFQTALKHE--CASSNDEILE--LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHR 132
                 : |  :: ::   :|:|: ||:  |:    |     :||   |: ||||:||:|
Db         64 -STALFPYSIHYQPIISSTNEWILQNILSLKKGD-----LCVAEYQTAGRGRRGRQWLSP 117

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192
                :|||| | || |:  :  || |  :|    |:      |:||||::    |||||
Db        118 FAGQIMFSFYWAFD-PKKSIEGLSLVIGLAIAEVLN-----VQVKWPNDILFDERKLGGI 171

Qy        193 LIETV-RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 251
              |:|      |    |:|||||  | |: | :   |   :      : :   ||  |:  |
Db        172 LVEIANHKNGMLNLVIGIGINVSLSKQTEIS---QPYAEVCEIDPDVERQTLLPKLIQHL 228

Qy        252 DAVLLQYARDGF-APFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAE 310
                |  : ::|    |   :|: |      : :| : : |  |  :|:| :| | :
Db        229 YTRLNIFEQNGIDEEFQQAWQSYNAFSNSEINVLTE-QGVISGIEQGIDERGYLKVLCGN 287

Qy        311 GKQTVVSGEISLR 323
                |    ||:|||
Db        288 KIQMFNGGEVSLR 300


RESULT 6
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             7.9%;  Score 243;  DB 9;  Length 249;
  Best Local Similarity   30.1%;  Pred. No. 1.4e-12;
  Matches   81;  Conservative   39;  Mismatches   89;  Indels   60;  Gaps   12;

Qy        341 LLLDGGNSRLKW-------AWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGC-- 391
              | || ||| :||       | :| |       |   ::: |          :||: ||
Db          3 LELDCGNSFIKWRVIHVADAVIEGGGIVDSDQALVAEVAAL---------ASVRLTGCRI 53
```

```
Qy        392 ----------AVCGEFKKA-QVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNA 440
                         |:|    :|  || ::|  :       :  |:|| |   :  | ||| |
Db         54 VSVRSEEETDALCALIAQAFAVQAKVAHPVR------EMAGVRNGYDDYQRLGMDRWLAA 107

Qy        441 LGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGK 500
              ||:   :: ||:|:  |||   | :: || :||| | ||  ||:  |   |  :
Db        108 LGAFHLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRI------ 161

Qy        501 RYPFPTTTGNAVAS-----GMMDAV-CGSVMMMHG----RLKE-KTGAGKPVDVIITGGG 549
              ||    :   |::|       ::||  | |:|: |    :|:: :   |:   | :|||
Db        162 RYD-DASAERALSSLSPGRSTVEAVERGCVLMLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        550 AAKVAEALPPAFLAENTVRVADNLVIHGL 578
              |  |  ||| |       ||  :||  ||
Db        221 APLVRAALPQA-------RVVPDLVFVGL 242


RESULT 7
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
  GENERAL INFORMATION:
   APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

  Query Match             7.5%;  Score 231.5;  DB 9;  Length 248;
  Best Local Similarity   27.7%;  Pred. No. 1.3e-11;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 397
              | || ||| :||  :|    :  |      |   :  | :: |      :  :  | :
Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56

Qy        398 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 450
              : :  : :||  :  | ||        |  |:|| |   :  | |||   : :   :: |
Db         57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116

Qy        451 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 503
              |:|:  |||||. | :  || :||| | ||  ||:  |   |       |   |      |
Db        117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176

Qy        504 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 558
                | | |  :     | ::|: | ::|:        |   :: :||| |  | : |
Db        177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228

Qy        559 PAFLAENTVRVADNLVIHGL 578
              |       |:  :||  ||
Db        229 GA-------RIMPDLVFVGL 241


RESULT 8
US-09-813-453A-65
; Sequence 65, Application US/09813453A
; Patent No. US20020168681A1
;  GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
;   LENGTH: 242
;   TYPE: PRT
;   ORGANISM: Xylessa fastidiosa
US-09-813-453A-65

  Query Match             6.4%;  Score 198.5;  DB 9;  Length 242;
  Best Local Similarity   30.7%;  Pred. No. 7.8e-09;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   23;  Gaps    8;

Qy        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401
              | | |||| | | :  |    |   ||   :        |   | | :  :|
Db          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63

Qy        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :|| ||
Db         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121

Qy        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518
              |:|:|  |  :| :||| |      ||:::|  |   |    |  .  |   | :|: ||
Db        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181

Qy        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 574
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   ||
Db        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228

Qy        575 IHGL 578
              : ||
Db        229 LDGL 232


RESULT 9
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             6.1%;  Score 187;  DB 9;  Length 249;
  Best Local Similarity   27.1%;  Pred. No. 7.6e-08;
  Matches   71;  Conservative   31;  Mismatches  114;  Indels   46;  Gaps    8;

Qy        341 LLLDGGNSRLKW-------------AWVENGT-----FATVGSAPYRDLSPLGAEWAEKV 382
              | || ||| :||               |:: |         :  | : |   :    ||:
Db          3 LELDCGNSFIKWRIITKSCSTLVSGGVVDSDTALLECLGNLSGAAFSDCRLVSVRSAEE- 61
```

```
Qy        383 DGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG 442
              ::| ||:   |  : |    | ::          |: | |      | |||   :|
Db         62 --TAKLV-CALADTFSISPVCAAPAPEL---------AGVINGYDDFARLGLDRWLAFVG 109

Qy        443 SRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN------R 496
              :    : ||:|:  |||:| | :   | :||| | ||  ||:  |   |  :
Db        110 AYHLVKGACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEAE 169

Qy        497 HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEA 556
               |  |      |  ||  |    : |  |      :|   |    : :||| |  ||:
Db        170 KALVRLVPGRATAEAVERGCSLMLRGFAMTQIEIAREYW--GDDFAIFVTGGDAVLVADV 227

Qy        557 LPPAFLAENTVRVADNLVIHGL 578
              || |       |:  :||  ||
Db        228 LPGA-------RIVPDLVFVGL 242


RESULT 10
US-09-738-626-4286
; Sequence 4286, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4286
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-4286

  Query Match             6.0%;  Score 183.5;  DB 9;  Length 288;
  Best Local Similarity   30.0%;  Pred. No. 1.8e-07;
  Matches   71;  Conservative   36;  Mismatches  113;  Indels   17;  Gaps    8;

Qy         92 SSNDEILELARI-APDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQY 150
              |:| ::| ||   ||:   ||:     |   ||| || ||   |   : |        |
Db         57 STNADLLALAGSGAPNWTVKTV---EFQDHARGRLGRPWSAPEGSQTIVSVLVQLSIDQV 113

Qy        151 E-LGSLSPVAAVACRRALSRLGLK-TQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVVG 208
              : :|::   | :|   ||: ||::   :|||||: :   || |||:|          |:|
Db        114 DRIGTIPLAAGLAVMDALNDLGVEGAGLKWPNDVQIHGKKLCGILVEATGFDSTPTVVIG 173

Qy        209 IGINFVLPKE---VENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAP 265
               | |  | ||   | :| |:      |      |   :|, |   | |: :
Db        174 WGTNISLTKEELPVPHATSL------ALEGVEVDRTTFLINMLTHLHTRLDQW-QGPSVD 226

Qy        266 FVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK-QTVVSGEIS 321
              :: :|:|     |: | :|  |:    |   ||   | : :  | |   |: :|||:
Db        227 WLDDYRAVCSSIGQDVRVLLPGDKELLGEAIGVATGGEIRVRDASGTVHTLNAGEIT 283


RESULT 11
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

  Query Match             5.9%;  Score 180.5;  DB 9;  Length 265;
  Best Local Similarity   27.0%;  Pred. No. 2.9e-07;
  Matches   67;  Conservative   35;  Mismatches  103;  Indels   43;  Gaps    8;

Qy        329 VSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRI 388
              :|   || ::   :|  |                :|  |     || |  : :|
Db         28 ISTDSRRTADELAVLLQG--------------LMGMHPL-----LGDELGDGID----- 62

Qy        389 VGCAVCGEFKKA--QVQEQLARKIEWLPSSAQALGIRNHY----RHPEEHGSDRWFNALG 442
               | |:|        :::|   |    :|:     |::        ||:| |:||  ||:
Db         63 -GIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPILTDHPKEVGADRIINAVA 121

Qy        443 SRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR--HAGK 500
              :        :||  ||| | ||::  | |:|| | ||  :  |:| |: | | :   |
Db        122 AVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQLRKIEVARP 181

Qy        501 RYPFPTTTGNAVASGMMDAVCGSVMMMHGRL-KEKTGAGKPVDVIITGGGAAKVAEALPP 559
              |       |  |: ||::    | |  :  |: :|       | || |||        | |
Db        182 RSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGG--------LAP 233

Qy        560 AFLAENTV 567
                | |::|
Db        234 MVLGESSV 241


RESULT 12
US-09-859-270-2
; Sequence 2, Application US/09859270
; Patent No. US20020048789A1
; GENERAL INFORMATION:
;  APPLICANT: Huang, Jianzhong
;  APPLICANT:  McDevitt, Damien
;  TITLE OF INVENTION: birA
;  FILE REFERENCE: GM10172
;  CURRENT APPLICATION NUMBER: US/09/859,270
;  CURRENT FILING DATE:  2001-05-17
;  PRIOR APPLICATION NUMBER: 09/175,172
;  PRIOR FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 311
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-859-270-2

  Query Match             5.8%;  Score 179;  DB 10;  Length 311;
  Best Local Similarity   28.0%;  Pred. No. 4.9e-07;
  Matches   75;  Conservative   45;  Mismatches  102;  Indels   46;  Gaps   12;
```

```
Qy          76 LGERSGFQTALKHECASSNDEILELARIAPDKAHK--TICVTHLQSKGRGRQGRKWSHRL 133
               | |    : : | |  |:  :    |: | |  |:  |: :   |: ||||  | :
Db          69 LEENLPIKVSFKPETKSTQLD----AKEAIDLGHEANTLYLASYQTAGRGRFQRSFYSPQ 124

Qy         134 GECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-KTQIKWPNDLVVGRDKLGGI 192
               |   | :     : |  :| | : : | |  :|:  | |    ||| ||: :   |:|||
Db         125 GGIYM-TLHLKPNLPYDKLPSYTLLVAGAVYKAIKNLTLIDVDIKWVNDIYLNNHKIGGI 183

Qy         193 LIE---TVRTGGKTVAVVGIGINFVL---PKEVENAASVQSLFQTASRRGNADAAVLLET 246
               | |   :| ||  |  ::|:|||| :   |:|::  |:  |||:       | | :
Db         184 LTEAMTSVETGLVTDIIIGVGINFTIKDFPQELKEKAA--SLFK-------ATAPITRNE 234

Qy         247 LLAEL---------DAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKG 297
               |: |:         : :|  | :  |              || |    : :  ::|  |
Db         235 LIIEIWRAFFETPAEELLYLYKKQSFI-----------LGKEVTFTLE-QKDYKGLAKD 281

Qy         298 VDGQGVLHLETAEGKQT-VVSGEISLRS 324
               :  | | ::   ||:  : ||||||
Db         282 ISENGKLLVQCDNGKEIWLNSGEISLNS 309


  SULT 13
S-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-813-453A-51

  Query Match             5.7%;  Score 175;  DB 9;  Length 260;
  Best Local Similarity   32.3%;  Pred. No. 8.3e-07;
  Matches   53;  Conservative   22;  Mismatches   81;  Indels    8;  Gaps    3;

          420 LGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG 479
               |||       | | |:||  ||:|:          ||:  ||| | | :  || : || | ||
Db           94 LGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPG 153

Qy         480 FHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAV-CGSVMMMHGRL-KEKTGA 537
                :|  ::|    | | | | :|       |    | |   |  | : :: | : : |
Db         154 INLSMQALHEAAAKLPRIAIQRPAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAER 213

Qy         538 GKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNL 581
               |:|: || ||| |:    |       ::      :| | ||| :
Db         214 GEPMTVIATGGVASLFEGA------TDSIDHFDSDLTIRGLLEI 251


RESULT 14
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

  Query Match             5.5%;  Score 168;  DB 9;  Length 250;
  Best Local Similarity   24.7%;  Pred. No. 3e-06;
  Matches   60;  Conservative   42;  Mismatches   97;  Indels   44;  Gaps    8;

Qy         341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVD--GNVRIVGCAVCGEFK 398
               |:|| ||: :             :|   | | : | |||    |   :    | |   |:
Db          20 LVLDVGNTNI----------VLGI--YND-TKLTAEWRLSTDVLRSADEYGIQVMNLFQ 65

Qy         399 KAQVQEQLARKI---EWLPSSAQAL--------------------GIRNHYRHPEEHGS 434
               : ::   |   :     :|:   :|                    ||   | :|:| |:
Db          66 QDKLDPTLVEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGA 125

Qy         435 DRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANL 494
               ||  ||: :    : : :::  ||| |  |: ::| |||| | ||  :  |:|  : | |
Db         126 DRIVNAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKL 185

Qy         495 NR--HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGK--PVDVIITGGGA 550
                |       |     | ::: ||::      |  :  :|||    |:     :::  ||
Db         186 PRVELIKPAYAICKNTISSIQSGIVYRYLRQVKYLFEKLKENLPDGRRTRTSLVLATGGL 245

Qy         551 AKV 553
               ||:
Db         246 AKL 248


RESULT 15
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
US-09-813-453A-45

  Query Match             5.5%;  Score 168;  DB 9;  Length 262;
  Best Local Similarity   31.2%;  Pred. No. 3.3e-06;
  Matches   53;  Conservative   25;  Mismatches   68;  Indels   24;  Gaps    5;

Qy         421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
               |:   | :| | |:||  ||:       :  ::|  ||| |   : :: ||:|| | ||
Db          95 GLNIKYENPREVGADRIVNAVAGIHLYGSPLIIVDFGTATTYCYINEEKHYMGGVITPGI 154

Qy         481 HLMKESLAVRTANLNR--------HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLK 532
                :  |:|  | | | |          ||       | :|: ||::    | |  :  |:|
Db         155 MISAEALYSRAAKLPRIEITKPSSVVGK------NTVSAMQSGILYGYVGQVEGIVKRMK 208
```

```
Qy        533 EKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADN-LVIHGLLNL 581
              |:  | :   || ||| |  ::|         | : | |  | : ||  |
Qb        209 EE--AKQEPKVIATGGLAKLISE-------ESNVIDVVDPFLTLKGLYML 249
```

Search completed: June 24, 2003, 22:24:06
Job time : 27.9909 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 17.2942 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-22
Perfect score: 3080
Sequence: 1 MTVLKPSHWRVLAELADGLP.....LVIHGLLNLIAAEGGESEHT 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	182.5	5.9	323	4	US-09-237-119-2	Sequence 2, Appli
2	179	5.8	311	4	US-09-175-172-2	Sequence 2, Appli
3	154	5.0	328	4	US-09-134-001C-3229	Sequence 3229, Ap
4	126	4.1	5588	4	US-09-036-987A-6	Sequence 6, Appli
5	126	4.1	5588	4	US-09-370-700-6	Sequence 6, Appli
6	122.5	4.0	4928	4	US-09-036-987A-5	Sequence 5, Appli
7	122.5	4.0	4928	4	US-09-370-700-5	Sequence 5, Appli
8	117.5	3.8	5087	4	US-09-144-085-1	Sequence 1, Appli
9	115	3.7	3816	4	US-09-428-517-3	Sequence 3, Appli
10	110.5	3.6	6095	4	US-09-144-085-2	Sequence 2, Appli
11	105	3.4	1222	2	US-08-682-517-15	Sequence 15, Appl
12	105	3.4	1252	2	US-08-682-517-9	Sequence 9, Appli
13	105	3.4	1891	2	US-08-804-227C-12	Sequence 12, Appl
14	105	3.4	1891	2	US-08-804-198-6	Sequence 6, Appli
15	104.5	3.4	4613	4	US-09-105-537-31	Sequence 31, Appl
16	104.5	3.4	11877	4	US-09-105-537-6	Sequence 6, Appli
17	103.5	3.4	3170	2	US-07-642-734C-5	Sequence 5, Appli
18	103.5	3.4	3170	3	US-08-439-009A-5	Sequence 5, Appli
19	100	3.2	444	6	5310667-5	Patent No. 5310667
20	98.5	3.2	4551	3	US-09-320-878-1	Sequence 1, Appli
21	97.5	3.2	674	4	US-09-199-637A-51	Sequence 51, Appl
22	96.5	3.1	357	4	US-09-605-785-826	Sequence 826, App
23	96.5	3.1	1228	4	US-09-605-785-537	Sequence 537, App
24	96.5	3.1	1228	4	US-09-439-313-537	Sequence 537, App
25	96.5	3.1	1261	4	US-09-605-785-538	Sequence 538, App
26	96.5	3.1	1261	4	US-09-439-313-538	Sequence 538, App
27	95.5	3.1	5215	4	US-09-105-537-2	Sequence 2, Appli
28	95	3.1	2475	4	US-09-413-814-48	Sequence 48, Appl
29	93.5	3.0	444	4	US-09-000-062-5	Sequence 5, Appli
30	93.5	3.0	570	3	US-08-832-078-5	Sequence 5, Appli
31	93.5	3.0	4550	2	US-08-804-227C-8	Sequence 8, Appli
32	93.5	3.0	4550	2	US-08-804-198-2	Sequence 2, Appli
33	93	3.0	1611	2	US-08-804-227C-5	Sequence 5, Appli
34	92.5	3.0	444	1	US-08-476-008-56	Sequence 56, Appl
35	92.5	3.0	444	1	US-08-306-063-56	Sequence 56, Appl
36	92.5	3.0	444	1	US-08-833-485-56	Sequence 56, Appl
37	92.5	3.0	444	4	US-09-243-374-5	Sequence 5, Appli
38	92.5	3.0	444	4	US-09-137-440-56	Sequence 56, Appl
39	92.5	3.0	444	4	US-09-000-062-3	Sequence 3, Appli
40	92.5	3.0	591	4	US-09-199-290-7	Sequence 7, Appli
41	92.5	3.0	756	4	US-09-085-199B-9	Sequence 9, Appli
42	92.5	3.0	4150	4	US-09-428-517-2	Sequence 2, Appli
43	92	3.0	852	4	US-09-741-150-2	Sequence 2, Appli
44	92	3.0	1525	4	US-09-396-651B-1	Sequence 1, Appli
45	91.5	3.0	618	4	US-09-199-290-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-09-237-119-2
; Sequence 2, Application US/09237119
; Patent No. 6448038
; GENERAL INFORMATION:
; APPLICANT: Kallender, Howard
; APPLICANT: Sylvester, Daniel R.
; APPLICANT: McDevitt, Damien
; APPLICANT: Huang, Jianzhong
; APPLICANT: Van Horn, Stephanie
; TITLE OF INVENTION: birA
; FILE REFERENCE: GM10189
; CURRENT APPLICATION NUMBER: US/09/237,119
; CURRENT FILING DATE: 1999-01-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-237-119-2

Query Match 5.9%; Score 182.5; DB 4; Length 323;
Best Local Similarity 26.0%; Pred. No. 1.4e-11;
Matches 78; Conservative 44; Mismatches 125; Indels 53; Gaps 14;

```
Qy         35 QQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSN 94
              |||   | |       |:: |:       :  |:||   : :  : :  | | |     :
Db         61 QQLPDIWYQ------GIIDQY------TKSSALFDFSEVYDSIDST--QLAAKKSLVGNQ 106

Qy         95 DEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGS 154
                    |  |:           |:|||||  | ||   |: |  |   |  ||
Db        107 SSFF----ILSDE----------QTKGRGRFNRHWSSSKGQGLWMS---VVLRPNVAFSM 149

Qy        155 LSP---VAAVACRRALSRLGL-KTQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV-GI 209
              :|      |:  | |:      : ::|||||: :   |: | | | |       |:: ||
Db        150 ISKFNLFIALGIRDAIQHFSQDEVKVKWPNDIYIDNGKVCGFLTEMVANNDGIEAIICGI 209

Qy        210 GINFVLPKEVEN--------AASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARD 261
              |||  | :::||        | |:|        :   |    || || |::    |:
Db        210 GIN--LTQQLENFDESIRHRATSIQ-----LHDKNKLDRYQFLERLLQEIEKRYNQFLTL 262

Qy        262 GFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK-QTVVSGEI 320
               |:    || ||:   :  ||  : :  |:|   :|  | | :    |:   ::| :|
Db        263 PFSEIREEYIAASNIWNR-TLLFTENDKQFKGQAIDLDYDGYLIVRDEAGESHRLISADI 321
```

RESULT 2
US-09-175-172-2
; Sequence 2, Application US/09175172

```
; Patent No. 6255075
; GENERAL INFORMATION:
;  APPLICANT: Huang, Jianzhong
;  APPLICANT: McDevitt, Damien
;  TITLE OF INVENTION: birA
;  FILE REFERENCE: GM10172
;  CURRENT APPLICATION NUMBER: US/09/175,172
;  CURRENT FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 311
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-175-172-2

  Query Match             5.8%;  Score 179;  DB 4;  Length 311;
  Best Local Similarity   28.0%;  Pred. No. 3.3e-11;
  Matches   75;  Conservative   45;  Mismatches  102;  Indels   46;  Gaps   12;

Qy         76 LGERSGFQTALKHECASSNDEILELARIAPDKAHK--TICVTHLQSKGRGRQGRKWSHRL 133
              | |    : : | |  |:  :    |: | |  |:  |: :   |: ||||  | :
Db         69 LEENLPIKVSFKPETKSTQLD----AKEAIDLGHEANTLYLASYQTAGRGRFQRSFYSPQ 124

Qy        134 GECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-KTQIKWPNDLVVGRDKLGGI 192
              |   | :     : |  :| | : : | |  :|:  | |    ||| ||: :    |:||
Db        125 GGIYM-TLHLKPNLPYDKLPSYTLLVAGAVYKAIKNLTLIDVDIKWVNDIYLNNHKIGGI 183

Qy        193 LIE---TVRTGGKTVAVVGIGINFVL---PKEVENAASVQSLFQTASRRGNADAAVLLET 246
              | |   :| ||  |  ::|:|||| :   |:|::  |:  |||:       | | :
Db        184 LTEAMTSVETGLVTDIIIGVGINFTIKDFPQELKEKAA--SLFK-------ATAPITRNE 234

Qy        247 LLAEL---------DAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKG 297
              |: |:         : :|  | :  |              || |    : :  ::|  |
Db        235 LIIEIWRAFFETPAEELLYLYKKQSFI-----------LGKEVTFTLE-QKDYKGLAKD 281

Qy        298 VDGQGVLHLETAEGKQT-VVSGEISLRS 324
              :   | | ::   ||:  : ||||||
Db        282 ISENGKLLVQCDNGKEIWLNSGEISLNS 309


RESULT 3
US-09-134-001C-3229
; Sequence 3229, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
   FILE REFERENCE: GTC-007
   CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3229
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3229

  Query Match             5.0%;  Score 154;  DB 4;  Length 328;
  Best Local Similarity   24.8%;  Pred. No. 2.2e-08;
  Matches   74;  Conservative   45;  Mismatches  141;  Indels   38;  Gaps   11;

Qy         33 KPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECAS 92
              |  |||    |              |  :|:|:       | |||  :  :  : |   |
Db         61 KGHQLNSLPDQW------------YSGIVKPI-------LDELGLFNHLE--VYHTVDS 98

Qy         93 SNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYEL 152
              :    |:  |         : ::  |::||||  | |    |: |  |     | |   :

Db         99 TQ---LKAKRALVGNKDTFLILSDEQTEGRGRFNRNWESSKGKGLWMSLVLRPDVPFSMI 155

Qy        153 GSLSPVAAVACRRALSRL-GLKTQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVV-GIG 210
                :   |:  | |: :    :  ||||||: :|  |: | | | |    :  |:: |||
Db        156 PKFNLFIALGIRDAIQQFSNERVTIKWPNDIYIGNKKICGFLTEMVANYDEIEAIICGIG 215

Qy        211 INFVLPKEVENAASVQSLFQTASRRGNADAAV----LLETLLAELDAVLLQYARDGFAPF 266
              ||      ||:        :  | | ::|: :     |  || ::     |:    |
Db        216 INM---NHVESDFDEDIKDKATSIRMHSDSIINRYTFLTALLTQIIHRFDQFLHQTFESI 272

Qy        267 VAEY-QAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK-QTVVSGEISL 322
                ||  | |  | : |   :    | |    :|  | | ::  :|:   ::| :| |
Db        273 REEYIHATNIWHRQ--LKFTENNHQFLGEAIDIDSDGFLIVKDEKGQLHRLMSADIDL 328


RESULT 4
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
;  GENERAL INFORMATION:
;    APPLICANT:  Baltz, Richard H.
;    APPLICANT:  Broughton, Mary C.
;    APPLICANT:  Crawford, Kathryn P.
;    APPLICANT:  Madduri, Krishnamurthy
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match             4.1%;  Score 126;  DB 4;  Length 5588;
  Best Local Similarity   18.8%;  Pred. No. 0.0043;
  Matches  153;  Conservative   79;  Mismatches  265;  Indels  318;  Gaps   32;

Qy         10 RVLAELADGLPQHVSQLARMAD------MKPQQLNGFWQQMPAHIRGLLRQHDGYW---- 59
              || : |   ||       |: |       :| ::  |:  :         |    ||
Db        747 RVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAEI-AFYSTVTGERTDTGRLDADYWYQNL 805

Qy         60 ----RLVRPLAVFDAEGLRELGERS-------GFQTALKHECASSNDEILELARIAPDKA 108
                  |  : :|    :| |   | |       | |  |  | | :   ::   |
Db        806 RQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGIQETL--EAADAGGVVVGSLR------ 857
```

```
Qy        109 HKTICVTHLQSKGRGRQG-RKWSHRLGEC----LMFSFGWVF------------------ 145
                                ||  | |:|   | ||    |  ::  ||
Db        858 ------------RGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGARRVPLPTYPFQRQ 904

Qy        146 ----DRPQYELGSLSPVAAVACRRAL---------------SRLGLKTQIKWPNDLVVG 185
                  :  :|: | |  |  ::    |                  :| :|||  |  | |||
Db        905 RYWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQ-PWLADHVVG 963

Qy        186 -----------------RDKLGGILIETV-----------------------RTGGKTV 204
                              |::|  ||| :                       | ::|
Db        964 GAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSV 1023

Qy        205 AV---------------------------VGIGINFVLPKEVENAASVQSLFQTASRRG 236
                |                          | | : : |     |  : | :  |: ||
Db       1024 RVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPLDSFYARAAERG 1083

Qy        237 ------------------NADAAVLLETLLAE-----------LDAVLLQYAR------D 261
                               | | | |   |           |||  |           :
 b       1084 FDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQALGAAEEDPDE 1143

Qy        262 GFAPFVAEYQAANRDHGKA----VLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
              |: ||   :|  :      |    | |:  |             || || ::      ::|
Db       1144 GWLPFA--WQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSID------SLVL 1195

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLD------GGNSRLKWAWVENGTFA-TVGSAPYRD 370
               :||    |: ::  :  : |    :|      |    : || : |   |   |||
Db       1196 RQIS----DKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSA---- 1247

Qy        371 LSPLGAEWAEKVDG------NVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRN 424
                || |    : |      :| :| |                          :| | | : :
Db       1248 ---LGTELHPDLTGLADPPPDVVVVPCG---------------------ASRQDLDVAS 1282

Qy        425 HYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTA------------------------ 459
                |   :   |   : | : ||: :  |||:|| |
Db       1283 EARAATQRMLDLIQDWLAAARFAGSRLVVVTCGAASTGPAEGVSDLVHAASWGLLRSAQS 1342

Qy        460 --------VTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN--------RHAGKRYP 503
                    | ||   :    |   :  |    :  ||:|   :                | |
Db       1343 ENPDRFVLVDVDGTAESWRALAAAVRSG----EPQLALRAGEVRVPRLARCVAAEDSRIP 1398

Qy        504 FPTTTGNAVASGMMDAVCGSV-----------MMMHGRLK-EKTGAGKPVDVIITGGGA 550
                |   |  : ||    : | |           ::: ||      |    || ::  | |
Db       1399 VPGADGTVLISGGTGLLGGLVARHLVAERGVRRLVLAGRRGWSAPGVTDLVDELVGLGAA 1458

Qy        551 AKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAE 585
               :||          :  | |   :  ||  |:||
Db       1459 VEVA----------SCDVGDRAQLDRLLTTISAE 1482
```

RESULT 5
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 5588
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

Query Match 4.1%; Score 126; DB 4; Length 5588;
Best Local Similarity 18.8%; Pred. No. 0.0043;
Matches 153; Conservative 79; Mismatches 265; Indels 318; Gaps 32;

```
Qy         10 RVLAELADGLPQHVSQLARMAD------MKPQQLNGFWQQMPAHIRGLLRQHDGYW---- 59
              ||  : |   ||       |: |       :| ::  |:  :         |    ||
Db        747 RVGIDYASHSPQIAQVQGRLLDRLGEVGSEPAEI-AFYSTVTGERTDTGRLDADYWYQNL 805

Qy         60 ----RLVRPLAVFDAEGLRELGERS-------GFQTALKHECASSNDEILELARIAPDKA 108
                  |  : :|    :| |   | |       | |  |  | | :   ::   |
Db        806 RQPVRFQQTVARMADQGYRFFVEVSPHPLLTAGIQETL--EAADAGGVVVGSLR------ 857

Qy        109 HKTICVTHLQSKGRGRQG-RKWSHRLGEC----LMFSFGWVF------------------ 145
                              ||  | |:|   | ||    |  ::  ||
Db        858 ------------RGEGGSRRWLTSLAECQVRGLPVNWEQVFLNTGARRVPLPTYPFQRQ 904

Qy        146 ----DRPQYELGSLSPVAAVACRRAL---------------SRLGLKTQIKWPNDLVVG 185
                  :  :|: | |  |  ::    |                  :| :|||  |  | |||
Db        905 RYWLESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQ-PWLADHVVG 963

Qy        186 -----------------RDKLGGILIETV-----------------------RTGGKTV 204
                              |::|  ||| :                       | ::|
Db        964 GAILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSV 1023

Qy        205 AV---------------------------VGIGINFVLPKEVENAASVQSLFQTASRRG 236
                |                          | | : : |     |  : | :  |: ||
Db       1024 RVHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPLDSFYARAAERG 1083

Qy        237 ------------------NADAAVLLETLLAE-----------LDAVLLQYAR------D 261
                               | | | |   |           |||  |           :
Db       1084 FDFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQALGAAEEDPDE 1143

Qy        262 GFAPFVAEYQAANRDHGKA----VLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
              |: ||   :|  :      |    | |:  |             || || ::      ::|
Db       1144 GWLPFA--WQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSID------SLVL 1195

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLD------GGNSRLKWAWVENGTFA-TVGSAPYRD 370
               :||    |: ::  :  : |    :|      |    : || : |   |   |||
Db       1196 RQIS----DKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSA---- 1247

Qy        371 LSPLGAEWAEKVDG------NVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRN 424
                || |    : |      :| :| |                          :| | | : :
Db       1248 ---LGTELHPDLTGLADPPPDVVVVPCG---------------------ASRQDLDVAS 1282

Qy        425 HYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTA------------------------ 459
                |   :   |   : | : ||: :  |||:|| |
Db       1283 EARAATQRMLDLIQDWLAAARFAGSRLVVVTCGAASTGPAEGVSDLVHAASWGLLRSAQS 1342

Qy        460 --------VTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN--------RHAGKRYP 503
                    | ||   :    |   :  |    :  ||:|   :                | |
Db       1343 ENPDRFVLVDVDGTAESWRALAAAVRSG----EPQLALRAGEVRVPRLARCVAAEDSRIP 1398

Qy        504 FPTTTGNAVASGMMDAVCGSV-----------MMMHGRLK-EKTGAGKPVDVIITGGGA 550
                |   |  : ||    : | |           ::: ||      |    || ::  | |
Db       1399 VPGADGTVLISGGTGLLGGLVARHLVAERGVRRLVLAGRRGWSAPGVTDLVDELVGLGAA 1458

Qy        551 AKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAE 585
               :||          :  | |   :  ||  |:||
Db       1459 VEVA----------SCDVGDRAQLDRLLTTISAE 1482
```

RESULT 6
US-09-036-987A-5
; Sequence 5, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy

```
;    APPLICANT:  Merlo, Donald J.
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4928 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-5

  Query Match             4.0%;  Score 122.5;  DB 4;  Length 4928;
  Best Local Similarity   21.6%;  Pred. No. 0.0086;
  Matches  182;  Conservative   75;  Mismatches  260;  Indels  325;  Gaps   44;

Qy         11 VLAELADGLPQ--------HVSQLARMADMKPQQLNGFWQ----------QMPAHIRGLL 52
              :| |:|| |           ||       |:|  |  | |          ::|  |  |: 
Db        745 LLTEIADRLHDQGCRTRWLRVSHAFHSPHMEP-MLEEFAQISRGREYHAPELPI-ISTLI 802

Qy         53 RQHDG--------YW-RLVRPLAVFDAEGLR----------ELGERSGFQTALKHECAS 92
               : ||         || | ||    | |||::          |||   |  : |  || :
Db        803 GELDGGRVMGTPEYWVRQVREPVRF-AEGVQALVGQGVGTIVELGP-DGALSTLVEECVA 860

Qy         93 SNDEI--LELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSF--------- 141
               :  :  : | |   |:| :|:     |   ||          ||    ||
Db        861 ESGRVAGIPLMRKDRDEA-RTVLAALAQIHTRG----------GEVDWRSFFAGTGAKQV 909

Qy        142 ---GWVFDRPQYELGS-----------------LSPVAAVACRRALSRLGLKT--QIK 177
                 : | | :| | |                 |  | |:|    :   |  |
Db        910 DLPTYAFQRQRYWLASTGRAGDVTAAGLAEADHPLLGAVVALADGEGVVLTGRLTAGSHP 969

Qy        178 WPND-LVVGRDKL-GGILIETV-----RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQ 230
               | :|  |:|   : |  ::| |     | |   |  : :    :||     |  || |
Db        970 WLSDHRVLGEIVVPGTAIVELVWHVGERLGCGRVEELALEAPLILPD--HGAVQVQVLVG 1027

Qy        231 TASRRGNADAA-----------------VLLETLLA-------------ELDAVLLQ 257
                 |   |                    |||  : |               |:||
Db       1028 PPGESGARSVALYSCPGEAIEPEWKKHATGVLLPPVAAENHELTAWPPENATEIDA---- 1083

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRD----GETVF------EGTVKGVDGQG----- 302
                ||   |:  :  |   :| |   ||     |  ||      :    |||  |
Db       1084 ---DGVYAFLEGHGFA---YGPAFRCLRGAWRRGGEVFAEVALPDDMQAGVDRFGVHPAL 1137

Qy        303 ---VLHLETAE--------------------GKQTVVSGEISLRSDD------------ 326
                  |||   ||                    :  ||   :|| |||
Db       1138 LDAVLHAAAAETSVVQSEARVPFSWRGVELRATESAVVRARLSLTSDDELSLVAVDPAGR 1197

Qy        327 ----------RPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPL-- 374
                       ||:|  : |          |:   :  |      | :|:    ||: :
Db       1198 FVATVDSLVTRPISRQQVRSGAI------GDCLFEVEWHRK---ALLGTTAGDDLAIVGD 1248

Qy        375 GAEWAEKVDGNVRIVGCAVCGEFKKA----------QVQEQLARKIEW--LPSSAQALG 421
               |  | | |    |    |   ||: |            |    | ::|   ||| ||
Db       1249 GPSWPESVRATARF---ATLDEFRAAVDSDVPAPGSVLVAAMSAEEVEGGSLPSRAQ--- 1302

Qy        422 IRNHYRHPEEHGSD------RWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 475
                      |  ||       |   |  ||: :  |||:   ||: |: :|    :| :
Db       1303 ---------ESTSDLLALVQSW---LADERFAESQLVVVT-RAAVSADSDSDVADLVGAS 1349

Qy        476 IM----------PGFHLMKE----------------------------SLAVRTANLNRH 497
                         ||  :: :                           :|  | | |
Db       1350 SWGLLSSAQSENPGRFVLVDVDGTPESWQALPAAVRAGEPQLALRRGVALVPRLARLTVR 1409

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSV-----------MMMHGR-----------LKEK 534
                    |  | |  : :|   |: | |             ::: ||           : |
Db       1410 EEGSSPQLDTDGTVLITGGTGALGGVVARHLVEEHGIRRLVLAGRRGWNAPGVHELVDEL 1469

Qy        535 TGAGKPVDVIITGGGAAKVAE---------ALPPAFLAENTVRVADNLVIHGLLNLIAAE 585
                ||  |:|:      |  ||:         |:|  :     |  | |    : :| ||:
Db       1470 ARAGAVVEVV-----ACDVADRTDLEHVLAAIPVDWPLRGIVHTAGVLADGVIGSLSAAD 1524

Qy        586 GG 587
               |
Db       1525 VG 1526


RESULT 7
US-09-370-700-5
; Sequence 5, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 4928
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-5

  Query Match             4.0%;  Score 122.5;  DB 4;  Length 4928;
  Best Local Similarity   21.6%;  Pred. No. 0.0086;
  Matches  182;  Conservative   75;  Mismatches  260;  Indels  325;  Gaps   44;

Qy         11 VLAELADGLPQ--------HVSQLARMADMKPQQLNGFWQ----------QMPAHIRGLL 52
              :| |:|| |           ||       |:|  |  | |          ::|  |  |: 
Db        745 LLTEIADRLHDQGCRTRWLRVSHAFHSPHMEP-MLEEFAQISRGREYHAPELPI-ISTLI 802

Qy         53 RQHDG--------YW-RLVRPLAVFDAEGLR----------ELGERSGFQTALKHECAS 92
               : ||         || | ||    | |||::          |||   |  : |  || :
Db        803 GELDGGRVMGTPEYWVRQVREPVRF-AEGVQALVGQGVGTIVELGP-DGALSTLVEECVA 860

Qy         93 SNDEI--LELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSF--------- 141
               :  :  : | |   |:| :|:     |   ||          ||    ||
Db        861 ESGRVAGIPLMRKDRDEA-RTVLAALAQIHTRG----------GEVDWRSFFAGTGAKQV 909

Qy        142 ---GWVFDRPQYELGS-----------------LSPVAAVACRRALSRLGLKT--QIK 177
```

```
Db        910 DLPTYAFQRQRYWLASTGRAGDVTAAGLAEADHPLLGAVVALADGEGVVLTGRLTAGSHP 969

Qy        178 WPND-LVVGRDKL-GGILIETV-----RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQ 230

Db        970 WLSDHRVLGEIVVPGTAIVELVWHVGERLGCGRVEELALEAPLILPD--HGAVQVQVLVG 1027

Qy        231 TASRRGNADAA----------VLLETLLA----------ELDAVLLQ 257

Db       1028 PPGESGARSVALYSCPGEAIEPEWKKHATGVLLPPVAAENHELTAWPPENATEIDA---- 1083

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRD----GETVF------EGTVKGVDGQG----- 302

Db       1084 ---DGVYAFLEGHGFA---YGPAFRCLRGAWRRGGEVFAEVALPDDMQAGVDRFGVHPAL 1137

Qy        303 ---VLHLETAE----------GKQTVVSGEISLRSDD---------- 326

Db       1138 LDAVLHAAAAETSVVQSEARVPFSWRGVELRATESAVVRARLSLTSDDELSLVAVDPAGR 1197

 y        327 ----------RPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPL-- 374

Db       1198 FVATVDSLVTRPISRQQVRSGAI------GDCLFEVEWHRK---ALLGTTAGDDLAIVGD 1248

Qy        375 GAEWAEKVDGNVRIVGCAVCGEFKKA----------QVQEQLARKIEW--LPSSAQALG 421

Db       1249 GPSWPESVRATARF---ATLDEFRAAVDSDVPAPGSVLVAAMSAEEVEGGSLPSRAQ--- 1302

Qy        422 IRNHYRHPEEHGSD------RWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 475

Db       1303 ---------ESTSDLLALVQSW---LADERFAESQLVVVT-RAAVSADSDSDVADLVGAS 1349

Qy        476 IM----------PGFHLMKE----------SLAVRTANLNRH 497

Db       1350 SWGLLSSAQSENPGRFVLVDVDGTPESWQALPAAVRAGEPQLALRRGVALVPRLARLTVR 1409

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSV----------MMMHGR----------LKEK 534

Db       1410 EEGSSPQLDTDGTVLITGGTGALGGVVARHLVEEHGIRRLVLAGRRGWNAPGVHELVDEL 1469

Qy        535 TGAGKPVDVIITGGGAAKVAE---------ALPPAFLAENTVRVADNLVIHGLLNLIAAE 585

Db       1470 ARAGAVVEVV-----ACDVADRTDLEHVLAAIPVDWPLRGIVHTAGVLADGVIGSLSAAD 1524

Qy        586 GG 587

Db       1525 VG 1526
```

ESULT 8
JS-09-144-085-1
; Sequence 1, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Claes
; APPLICANT: Betlach, Mary C.
; APPLICANT: Ashley, Gary
; APPLICANT: Julien, Bryan
; APPLICANT: Ziermann, Rainer
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 30062-20020.20
; CURRENT APPLICATION NUMBER: US/09/144,085
; CURRENT FILING DATE: 1998-08-31
; EARLIER APPLICATION NUMBER: 09/010,809
; EARLIER FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 5087
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-144-085-1

```
  Query Match             3.8%;  Score 117.5;  DB 4;  Length 5087;
  Best Local Similarity   20.0%;  Pred. No. 0.033;
  Matches  135;  Conservative   84;  Mismatches  252;  Indels  205;  Gaps   32;

Qy         34 PQQLNGFWQQMPAHIRGLLRQHDGYWRL-------VRPLAV--FDAEGLRELGERSGFQT 84

Db       3240 PIAIVGVGLRMPGGANDL----ESFWQVLVEGRDTLRPIPTDRFDVEAMYDPNPEAKGKT 3295

Qy         85 ALKHECASSNDEILELAR----IAP------DKAHKTICVTHL----------QSKG-- 121

Db       3296 YVKH--ASLLDDVASFDAGFFGISPREAEPMDPQHRLLLETAWSALEDAGVRPDQLKGSD 3353

Qy        122 --------------RGRQGRKWSHRL-GECLMFSFGWVFDRPQYELGSLSPVAAV--AC 163

Db       3354 TGVFVGVAPSEYASYRGKSANEDAYALTGTALSFAAG----RVAYHLGLQGPAVSVDTAC 3409

Qy        164 RRALSRLGLKTQIKWPNDLVVG--------RDKLGGILIETVRTGG-----KTVAVV--- 207

Db       3410 SSSLVALHLACDALRRGDCEVALAAGVQVLANPAGFVLLSRTRAVSPDGRCKTFSQAADG 3469

Qy        208 ---GIGINFVLPKEVENAAS-----VQSLFQTASRRGNADAAVLLETLLAELDAVLLQYA 259

Db       3470 YGRGEGVGVVVLMRLSDAQAQGMRVLGVVRGTAVNQDGASSGITAPNGTAQQKVVRAALR 3529

Qy        260 RDGFAPFVAEYQAANRD----HGKAVLLLRDGETVFEGTVKGVDGQG------------- 302

Db       3530 NAGL-------EASSIDVVECHGTGTSL---GDPIEVQALGAVYGQGREATRPLRLGAVK 3579

Qy        303 --VLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Db       3580 SNIGHLESAAGIAGVCKILAAFRHEALPATL---HSSPR-------NPQISWE------- 3622

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCG-----------EFKKAQVQEQLARK 409

Db       3623 ----SLPVQVVDRL-TGWPRRADGLPRFAGVSSFGISGTNAHVILEEAPLEAVREPAAVR 3677

Qy        410 IEWLPSSAQALGIRNHYRHPEEHG--SDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD 467

Db       3678 -EPLAAEGVAIPLLLSGRDEASVGAQAERWAKWLGEHAEVRWPDV--------------- 3721

Qy        468 DGHYLGGTIMPGFHLMKESLAVRTANLNR-HAGKRYPFPTTTGNAVASGMMDAVCGSVMM 526

Db       3722 --------------------VRTAALHRTHFAWR-------ASVQAASVSEAVEGLRAL 3753

Qy        527 MHGR-----LKEKTGAGKPVDVIITGGGAAKVAEA-----LPPAFLA--ENTVRVADNLV 574

Db       3754 SEGRAAAGVVRGTGGRGGKLAVLFTGQGSQRLGMGKRLYEVYPVFRAAFDEVCEALDAHL 3813

Qy        575 IHGLLNLIAAEGGESE 590

Db       3814 DRGLREVVFAEAGSEQ 3829
```

RESULT 9
US-09-428-517-3
; Sequence 3, Application US/09428517
; Patent No. 6251636
; GENERAL INFORMATION:
; APPLICANT: Betlach, Mary C.
; APPLICANT: Shah, Sanjay Krishnakant
; APPLICANT: McDaniel, Robert
; APPLICANT: Tang, Li
; TITLE OF INVENTION: RECOMBINANT OLEANDOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062-20029.00
; CURRENT APPLICATION NUMBER: US/09/428,517
; CURRENT FILING DATE: 1999-10-28
; EARLIER APPLICATION NUMBER: 60/120,254
; EARLIER FILING DATE: 1999-02-16
; EARLIER APPLICATION NUMBER: 60/106,100
; EARLIER FILING DATE: 1998-10-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3

```
;   LENGTH: 3816
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Recombinant
;   OTHER INFORMATION: Oleandolide PKS
US-09-428-517-3

  Query Match             3.7%;  Score 115;  DB 4;  Length 3816;
  Best Local Similarity   20.2%;  Pred. No. 0.038;
  Matches  116;  Conservative   75;  Mismatches  205;  Indels  178;  Gaps   28;

Qy         96 EILELARIAPDKAHKTICVTHLQSKGR--------------GRQGRKWSHRLGECLMFSF 141
              |::| | :|||  |::   | : | |:              | ||      :|
Db       1693 ELIESAGVAPDSLHRSRTGTFIGSNGQFYAPLLWNSGGDLEGYQG------VGNAGSVMS 1746

Qy        142 GWVFDRPQYELGSLSPVAAV--ACRRALSRLGLKTQ-----------------IKWPNDL 182
              |    | | ||   |   | ||  :|  | | |                   :  |:
Db       1747 G----RVAYSLGLEGPAVTVDTACSSSLVALHLAVQALRRGECSLAIAGGVTVMSTPDSF 1802

Qy        183 VV----------GRDK--------------LGGILIETV----RTGGKTVAVV-GIGINF 213
              |           || |              :  :|:| :    | | : :||| |  :|
Db       1803 VEFSRQQGLSEDGRCKAFASTADGFGLAEGVSALLVERLSDARRLGHRVLAVVRGSAVN- 1861

Qy        214 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFA---PFVAEY 270
                  :: ||          :     | | :  ||  |  :::     |     |  |:
Db       1862 ------QDGASNGLTAPNGPSQQRVIRAALADAGLAPADVDVVEAHGTGTRLGDPIEAQA 1915

Qy        271 QAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDDRPVS 330
                |    |:|      |  |  |:||  |    | : | |   |:   ::|   :| |
Db       1916 LLATYGQGRA-----GGRPVVLGSVKSNIG----HTQAAAGVAGVMKMVLAL---ERGV- 1962

Qy        331 VPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVG 390
              |||       | :|  :  : |             ||   :|:     |:    | ||  |
Db       1963 VPKT------LHVDEPSPHVDW-----------SAGEVELAVEAVPWSR--GGRVRRAG 2002

Qy        391 CA---VCGEFKKAQVQEQLARK--------------IEWLPSSAQALGIRNHYRHPEEH 432
               :   : |       |:|  |                : |: |   |  :|        |
Db       2003 VSSFGISGTNAHVIVEEAPAEPEPEPGTRVVAAGDLVVPWVVSGRDAGALREQAARLAAH 2062

Qy        433 GS-------DRWFNALGSRRFSRNACVVVSCGTAVTVDALTDD-GHYLGGTIMPGFHLMK 484
              |       |  :: : :|    :  |:|       :|:: :    :  | ::||
Db       2063 VSSTGAGVVDVGWSLVATRSVFEHRAVMVGS----ELDSMAESLAGFAAGGVVPGV---- 2114

Qy        485 ESLAVRTANLNRHAGKR----YPFPTTTGNAVASGMMDAV---------CGSVMMMHGRL 531
                    : :    |:|    :|   :    :|:|::||           | :|:
Db       2115 ------VSGVAPAEGRRVVFVFPGQGSQWVGMAAGLLDACPVFAEAVAECAAVL------ 2162

Qy        532 KEKTGAGKPVDVIITGGGAAKV--AEALPPAFLA 563
                  |  :  :: ||| | :   : : ||  |
Db       2163 --DPVTGWSLVEVLRGGGEAVLGRVDVVQPALWA 2194


RESULT 10
US-09-144-085-2
; Sequence 2, Application US/09144085
; Patent No. 6280999
; GENERAL INFORMATION:
;  APPLICANT: Gustafsson, Claes
;  APPLICANT: Betlach, Mary C.
;  APPLICANT: Ashley, Gary
;  APPLICANT: Julien, Bryan
;  APPLICANT: Ziermann, Rainer
;  TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA
;  TITLE OF INVENTION: THEREFOR
;  FILE REFERENCE: 30062-20020.20
;  CURRENT APPLICATION NUMBER: US/09/144,085
;  CURRENT FILING DATE: 1998-08-31
;  EARLIER APPLICATION NUMBER: 09/010,809
;  EARLIER FILING DATE: 1998-01-22
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2

;   LENGTH: 6095
;   TYPE: PRT
;   ORGANISM: Sorangium cellulosum
US-09-144-085-2

  Query Match             3.6%;  Score 110.5;  DB 4;  Length 6095;
  Best Local Similarity   20.5%;  Pred. No. 0.27;
  Matches  157;  Conservative   83;  Mismatches  266;  Indels  259;  Gaps   39;

Qy          8 HWRVLAELADGLPQHVSQLARMA--DMKPQQLNGFWQQMPAHIR--GLLRQHDGYWRLVR 63
              :||:|||  | : :  |:     :  |  |: :         :::|  | ||  |
Db       5156 YWRLLAEGKDAIERFPSRYDAFSVYDPDPEAVG------KSYVREGGFLRDID------- 5202

Qy         64 PLAVFDAEGLRELGERSGFQTALKHECASSND----------EILELARIAPDKAHKTIC 113
                 ||||          ||      |  : :           | || | : |    ::
Db       5203 ---VFDA----------GFFGISPREAQAMDPQQRLVLETAWEALERAGVRPSMLSESAT 5249

Qy        114 VTHLQSKGRGRQGRKWSHRLGECLMFSFGW---------VFDRPQYELGSLSPVAAV--A 162
                :|      |  |  :   ||  |    |:          :  |  | ||   |   |  |
Db       5250 GVYL-----GWMGSDYGALLGNDLAALDGYQGTGSAASVLSGRVAYVLGLQGPAITVDTA 5304

Qy        163 CRRALSRLGLK-TQIKWPN-DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVE 220
              |  :|  | |  | ::    || :      ||::: |   |             ||   |
Db       5305 CSSSLVSLHLACTALRQGECDLAL----TGGVMVMTTPAG-----------FV---EFS 5345

Qy        221 NAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPF-VAEYQAANRD--- 276
              |  :    :  |    ||  :  |     |   |    |||      |    | |:|
Db       5346 RARGLARDGRCKSFSAQADGVIWSEGCGMLLLKRLSDARRDGDRVLGVIRGSAVNQDGRS 5405

Qy        277 ---------------------------------HGKAVLLLRDGETVFEGTVKGVDG-- 300
                                                ||    |   |: :  | :  | |
Db       5406 QGLTAPNGPAQQRVIRQALSSCGLSPEDIDAVEAHGTGTSL---GDPIEAGALAEVFGPE 5462

Qy        301 ------------QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGN 347
                         : | : | |   |:   ::|: :  | ::   : |        |
Db       5463 RSPERPLYLGSSKSNLGHAQAAAGVAGVIKMVLALQHEVLPKTLHAEQPSPHIAWEGSGL 5522

Qy        348 SRLKWA--WVENG--------TFATVGSAPYRDLSPLGAE-WAEKVDGNV--RIVGCAVC 394
              | |: |  |  ||         :|   |:  :  |    ||   | |:      ::   :
Db       5523 SLLQEARPWRRNGRVRRAGVSSFGISGTNAHIILEEAPAEARREPVEAEAAPALLPLVLS 5582

Qy        395 GEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFN-----ALGSRRF-SR 448
              |   :| |  |  |  :||                 ||||   | :     ||    | ||
Db       5583 GR-DEASVAAQAGRWAKWL---------------EEHGEVGWSDVVRTAALHRTHFESR 5625

Qy        449 NACVVVSCGTAVTV-DALTDD-GH----------------YLG-GTIMPGF--HLMKES 486
              : :  |    | |  ||::  ||                : | |:  ||    |:::|
Db       5626 ASMLAASVSEVVEVLRALSEGRGHRAVSVGTARARGKVVFVFPGQGSQWPGMGRALLEQS 5685

Qy        487 LAVRTA-----------------NLNRHAGKRYP-----------FPTTTGNAVA--- 513
              |  |                    |    |:  |                |    | | |
Db       5686 AAFAEAVQACDEALRPWTGWSVLSVLRGDGGEEQPSLERVDVVQPALFAMCVGLAAAWRS 5745

Qy        514 -----------SGMMDA--VCGSVMMMHG---------RLKEKTGAGKPVDVIITGGGA 550
                          | : |  |||:: :  |           :::::| |    :::
Db       5746 LGLEPAAVVGHSQGEVSAAVVCGALSLAEGARVVALRSQAVRQQSGMGA---MMLVEQPV 5802

Qy        551 AKVAEALPP-----AFLAEN----TVRVADNLVIHGLLNLIAAEG 586
              ::| | : |     |  | |    ||   |   : ||:  : |||
Db       5803 SEVQERIAPYGEALAIAAVNTSNSTVVSGDVEAVDGLMVELTAEG 5847


RESULT 11
US-08-682-517-15
; Sequence 15, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
```

```
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1222 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-15

  Query Match             3.4%;  Score 105;  DB 2;  Length 1222;
  Best Local Similarity   21.1%;  Pred. No. 0.067;
  Matches  100;  Conservative   51;  Mismatches  168;  Indels  156;  Gaps   23;

Qy        194 IETVRTGGKTVAV----VGIGIN---------------FVLPKEVENA-ASVQSLFQTA 232
              :|     ||| ||     || :|               |    || |:   | : | ||
Db        736 LEVTSVDGKTTAVKVIATGIAVNTDGKDYAFTAKEATATFTATNEVPNSYTGVATQFNTA 795

Qy        233 SRRGNAD-----------AAVLLETLL-----------AELDAVLLQYARDGFAPFVA 268
                |::           | |  :|              |  :|:| ||| :|
Db        796 DSGSNSNSIWFAGKNPVKYAGVSGKTYKYFGANGNEVFGEAAWEALLTQYATEG-QKVTI 854

Qy        269 EYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQ----------------GVLHLETAEG 311
               |               ||:||    :  |:                    | | |  | |
Db        855 SYNV------------DGDTVTFKVISAVNSSTEAIKPVAPTTPAAPTTGALTLTPAAG 901

Qy        312 ---KQTVVSGEISLRSDDRPVSVPKRR-DSERFLLLDGGNSRLKWAWVE----NGTFATV 363
                  |  :  : :   |  ::|     |:    | |  |: |    ||     | |||
Db        902 GLVDLTTATNTLGISLADADLNVSATTVDTATVSLKDSANNSLSLTLVETGANTGVFATT 961

Qy        364 GSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIR 423
               |    || | |       | : :       :  |:    :|  |    :  :  |
Db        962 VQA--GTLSSLTA-------GTLTVT-------YADAKNAAGVAENITASVTLKKTTG-- 1003

Qy        424 NHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAV--TVDALTDDGHYLGGTIMPGFH 481
                           |: |  |::   |: |  ||   |  ::  |  :  |    ||
Db       1004 ---------------AITSDTFTQG--VLPSAATAAEYTSKSIAADYTFATG---EGFT 1042

Qy        482 LMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAV----CGSVMMMHGRLKEK--- 534
               |  ::   :  ||   |||:          | |: ||:     |:  :   ::  |
Db       1043 LNIDNAGAQVINL---AGKK----------GAQGVADAINATFAGTATVSGDKVVIKSAT 1089

Qy        535 TGAGKPVDV-----------IITGGGAAKVAEALPPAFLAENTVRVADNLVIHGL 578
               || |  |:|           :: |        |   ||     : | : :|| |:
Db       1090 TGVGSEVEVTFSSVNQVLNAVVNGKDQVVAGTAATKAFTITTALSVGEKVVIDGV 1144


RESULT 12
US-08-682-517-9
; Sequence 9, Application US/08682517
; Patent No. 5874267
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  Expression of surface layer proteins
;    NUMBER OF SEQUENCES:  25
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/682,517
;      FILING DATE:
;      CLASSIFICATION:
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1252 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-682-517-9

  Query Match             3.4%;  Score 105;  DB 2;  Length 1252;
  Best Local Similarity   21.1%;  Pred. No. 0.07;
  Matches  100;  Conservative   51;  Mismatches  168;  Indels  156;  Gaps   23;

Qy        194 IETVRTGGKTVAV----VGIGIN---------------FVLPKEVENA-ASVQSLFQTA 232
              :|     ||| ||     || :|               |    || |:   | : | ||
Db        766 LEVTSVDGKTTAVKVIATGIAVNTDGKDYAFTAKEATATFTATNEVPNSYTGVATQFNTA 825

Qy        233 SRRGNAD-----------AAVLLETLL-----------AELDAVLLQYARDGFAPFVA 268
                |::           | |  :|              |  :|:| ||| :|
Db        826 DSGSNSNSIWFAGKNPVKYAGVSGKTYKYFGANGNEVFGEAAWEALLTQYATEG-QKVTI 884

Qy        269 EYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQ----------------GVLHLETAEG 311
               |               ||:||    :  |:                    | | |  | |
Db        885 SYNV------------DGDTVTFKVISAVNSSTEAIKPVAPTTPAAPTTGALTLTPAAG 931

Qy        312 ---KQTVVSGEISLRSDDRPVSVPKRR-DSERFLLLDGGNSRLKWAWVE----NGTFATV 363
                  |  :  : :   |  ::|     |:    | |  |: |    ||     | |||
Db        932 GLVDLTTATNTLGISLADADLNVSATTVDTATVSLKDSANNSLSLTLVETGANTGVFATT 991

Qy        364 GSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIR 423
               |    || | |       | : :       :  |:    :|  |    :  :  |
Db        992 VQA--GTLSSLTA-------GTLTVT-------YADAKNAAGVAENITASVTLKKTTG-- 1033

Qy        424 NHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAV--TVDALTDDGHYLGGTIMPGFH 481
                           |: |  |::   |: |  ||   |  ::  |  :  |    ||
Db       1034 ---------------AITSDTFTQG--VLPSAATAAEYTSKSIAADYTFATG---EGFT 1072

Qy        482 LMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAV----CGSVMMMHGRLKEK--- 534
               |  ::   :  ||   |||:          | |: ||:     |:  :   ::  |
Db       1073 LNIDNAGAQVINL---AGKK----------GAQGVADAINATFAGTATVSGDKVVIKSAT 1119

Qy        535 TGAGKPVDV-----------IITGGGAAKVAEALPPAFLAENTVRVADNLVIHGL 578
               || |  |:|           :: |        |   ||     : | : :|| |:
Db       1120 TGVGSEVEVTFSSVNQVLNAVVNGKDQVVAGTAATKAFTITTALSVGEKVVIDGV 1174


RESULT 13
US-08-804-227C-12
; Sequence 12, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
```

```
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1891 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-227C-12

  Query Match             3.4%;  Score 105;  DB 2;  Length 1891;
  Best Local Similarity   22.2%;  Pred. No. 0.14;
  Matches  146;  Conservative   67;  Mismatches  246;  Indels  198;  Gaps   33;

Qy         32 MKPQQ---LNGFWQQM------PAHIRG------LLRQHDGYWRLVRPLAVFDAEGLREL 76
Db        121 MDPQQRLLLETSWETLERAGIVPASLRGSRTGVFVGAAHTGY--------VTDTARAPEG 172

Qy         77 GERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGEC 136
Db        173 TE--GYLLTGNADAVMSGRIAYSLGLEGPALTIGTACSSSLVALHLAVQ----SLRRGEC 226

Qy        137 LMFSFGWVFDRPQYELGSLSPVAAV--ACRRALSRLGLKTQIKWPNDLVVGRDKLGGILI 194
Db        227 DLALAGGVAVMP-------DPTVFVEFSRQRGLAVDGRCKAFAEGADGTAWAEGVGVLLV 279

Qy        195 ETV----RTGGKTVAVV-GIGINFVLPKEVENAASVQS--LFQTASRRGNADAAV----- 242
Db        280 ERLSDARRNGHRVLAVVRGSAVN---QDGASNGLTAPSGPAQQRVIREALADAGLTPADV 336

Qy        243 -LLE-----TLLA---ELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEG 293
Db        337 DVVEAHGTGTALGDPIEAGALLATYGRE----------------------RVGDPLWLG 373

Qy        294 TVKGVDGQGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWA 353
Db        374 SLKSNIG----HAQAAAG----VGGVIKVVQAMRHGSLP------RTLHVDAPSSKVEWA 419

Qy        354 WVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWL 413
Db        420 ---SGAV---------ELLTEGRSWPRRVERVRRAAVSAFGVSGTNAHV------VLEEA 461

Qy        414 PSSAQALGIRNHYRHPEEHGSDR---------WFNALGSRRFSRNACVVVSCGTAVTVDA 464
Db        462 PVEAGS-------EHGDGPGPDRPDAVTGPLPWVLSARSREALR--------GQAGRLAA 506

Qy        465 LTDDGHYLG-----GTIMP----GFHL--MKESL---AVRTANLNRHAGKRYPFPTT--T 508
Db        507 LARQGRTEGTGGGSGLVVPAADIGYSLATTRETLEHRAVALVQENRTAGEDLAALAAGRT 566

Qy        509 GNAVASGMMDAVCGSVMMMHGRLKEKTGAG-------------------------KPVD 542
Db        567 PESVVTGVARRGRGIAFLCSGQGAQRLGAGRELRGRFPVFADALDEIAAEFDAHLERPLL 626

Qy        543 VIITGGGAAKVAEAL------PPAFLAENTV---------RVADNLVIHGLLNLIAA 584
Db        627 SVMFAEPATPDAALLDRTDYTQPALFAVETALFRLLESWGLVPDVLVGHSIGGLVAA 683


RESULT 14
US-08-804-198-6
; Sequence 6, Application US/08804198
; Patent No. 5945320
;  GENERAL INFORMATION:
;    APPLICANT:  Burgett, Stanley G.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rao, Nagaraja R.
;    APPLICANT:  Richardson, Mark A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    TITLE OF INVENTION:  PLATENOLIDE SYNTHASE GENE
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PAUL R. CANTRELL  1138
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Macintosh
;      OPERATING SYSTEM:  Macintosh 7.0
;      SOFTWARE:  Microsoft Word 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,198
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  CANTRELL, PAUL R.
;      REGISTRATION NUMBER:  36,470
;      REFERENCE/DOCKET NUMBER:  P9113
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3885
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1891 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
US-08-804-198-6

  Query Match             3.4%;  Score 105;  DB 2;  Length 1891;
  Best Local Similarity   22.2%;  Pred. No. 0.14;
  Matches  146;  Conservative   67;  Mismatches  246;  Indels  198;  Gaps   33;

Qy         32 MKPQQ---LNGFWQQM------PAHIRG------LLRQHDGYWRLVRPLAVFDAEGLREL 76
Db        121 MDPQQRLLLETSWETLERAGIVPASLRGSRTGVFVGAAHTGY--------VTDTARAPEG 172

Qy         77 GERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGEC 136
Db        173 TE--GYLLTGNADAVMSGRIAYSLGLEGPALTIGTACSSSLVALHLAVQ----SLRRGEC 226

Qy        137 LMFSFGWVFDRPQYELGSLSPVAAV--ACRRALSRLGLKTQIKWPNDLVVGRDKLGGILI 194
Db        227 DLALAGGVAVMP-------DPTVFVEFSRQRGLAVDGRCKAFAEGADGTAWAEGVGVLLV 279

Qy        195 ETV----RTGGKTVAVV-GIGINFVLPKEVENAASVQS--LFQTASRRGNADAAV----- 242
Db        280 ERLSDARRNGHRVLAVVRGSAVN---QDGASNGLTAPSGPAQQRVIREALADAGLTPADV 336

Qy        243 -LLE-----TLLA---ELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEG 293
Db        337 DVVEAHGTGTALGDPIEAGALLATYGRE----------------------RVGDPLWLG 373

Qy        294 TVKGVDGQGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWA 353
Db        374 SLKSNIG----HAQAAAG----VGGVIKVVQAMRHGSLP------RTLHVDAPSSKVEWA 419

Qy        354 WVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWL 413
Db        420 ---SGAV---------ELLTEGRSWPRRVERVRRAAVSAFGVSGTNAHV------VLEEA 461

Qy        414 PSSAQALGIRNHYRHPEEHGSDR---------WFNALGSRRFSRNACVVVSCGTAVTVDA 464
Db        462 PVEAGS-------EHGDGPGPDRPDAVTGPLPWVLSARSREALR--------GQAGRLAA 506

Qy        465 LTDDGHYLG-----GTIMP----GFHL--MKESL---AVRTANLNRHAGKRYPFPTT--T 508
Db        507 LARQGRTEGTGGGSGLVVPAADIGYSLATTRETLEHRAVALVQENRTAGEDLAALAAGRT 566

Qy        509 GNAVASGMMDAVCGSVMMMHGRLKEKTGAG-------------------------KPVD 542
Db        567 PESVVTGVARRGRGIAFLCSGQGAQRLGAGRELRGRFPVFADALDEIAAEFDAHLERPLL 626

Qy        543 VIITGGGAAKVAEAL------PPAFLAENTV---------RVADNLVIHGLLNLIAA 584
Db        627 SVMFAEPATPDAALLDRTDYTQPALFAVETALFRLLESWGLVPDVLVGHSIGGLVAA 683
```

```
RESULT 15
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
   LENGTH: 4613
   TYPE: PRT
   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             3.4%;  Score 104.5;  DB 4;  Length 4613;
  Best Local Similarity   22.4%;  Pred. No. 0.79;
  Matches  104;  Conservative   52;  Mismatches  174;  Indels  135;  Gaps   22;

Qy        142 GWVFDRPQYELGSLSPVAAV------------AC---RRALSRLGLK--TQIKWPNDLV 183
              | : :|  | ||   |   |             ||   ||  | | |     :    | :
Db        237 GIIANRLSYTLGLRGPSMVVDSGQSSSLVAVHLACESLRRGESELALAGGVSLNLVPDSI 296

Qy        184 VGRDKLGGILIE------TVRTGGKTVAVVGIGINFVLPKEVENAAS----VQSLFQTAS 233
              :|  | ||:  :        |  |    | | |  ||: | :  | :    | :: :  |
Db        297 IGASKFGGLSPDGRAYTFDARANG---YVRGEGGGFVVLKRLSRAVADGDPVLAVIR-GS 352

Qy        234 RRGNADAAVLLETLLAELDAVLLQYA--RDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                |  ||  : |  |:    :|: |  | | ||    |      ||    :    |
Db        353 AVNNGGAAQGMTTPDAQAQEAVLREAHERAGTAPADVRYVEL---HGTGTPVGDPIEAAA 409

Qy        292 EGTVKGVD---GQGVL---------HLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSER 339
              |   |     || :|         ||| | |   ::   :::|    | |:
Db        410 LGAALGTGRPAGQPLLVGSVKTNIGHLEGAAGIAGLIKAVLAVRGRALPASL-------- 461

Qy        340 FLLLDGGNSRLKWAWVENGTFATVGSA-PYRDLS-PLGAE---WAEKVDGNVRIVGCAVC 394
                           : |   | |: :|:  :  |   |  : ||   :|| :
Db        462 ------------------NYETPNPAIPFEELNLRVNTEYLPWEPEHDGQRMVVGVSSF 502

 Y        395 GEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVV 454
              |                 :|  | :   ||        | |     | | ||
Db        503 G----------------------MGGTNAHVVLEE--------APGG---CRGASVVE 527

Qy        455 SCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHA--GKRYPFPTTTGNAV 512
              |          |  |  :|| ::| : :  :| |   | : | | .  |          ||
Db        528 S----------TVGGSAVGGGVVP-WVVSAKSAAALDAQIERLAAFASRDRTDGVDAGAV 576

Qy        513 ASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
               :| :||   : ::  || : :  |      :: | |   :| ||
Db        577 DAGAVDAGAVARVLAGGRAQFEHRA------VVVGSGPDDLAAAL 615


Search completed: June 24, 2003, 21:49:38
Job time : 21.2942 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 48.2837 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-22
Perfect score: 3080
Sequence: 1 MTVLKPSHWRVLAELADGLP.....LVIHGLLNLIAAEGGESEHT 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3080	100.0	592	16	Q9JWI7	Q9jwi7 neisseria m
2	3024	98.2	592	16	Q9JXF1	Q9jxf1 neisseria m
3	326	10.6	291	16	Q8Y2M3	Q8y2m3 ralstonia s
4	305.5	9.9	321	16	Q8X709	Q8x709 escherichia
5	295	9.6	288	2	Q9ZIU6	Q9ziu6 bordetella
6	287.5	9.3	320	16	Q9KV39	Q9kv39 vibrio chol
7	281.5	9.1	312	16	Q9HWC0	Q9hwc0 pseudomonas
8	279.5	9.1	295	16	Q8Y2M4	Q8y2m4 ralstonia s
9	276.5	9.0	319	16	Q8ZAN5	Q8zan5 yersinia pe
10	269.5	8.8	320	16	Q8Z317	Q8z317 salmonella
11	246	8.0	326	16	Q8R8N0	Q8r8n0 thermoanaer
12	244	7.9	322	2	O32511	O32511 deinococcus
13	244	7.9	474	16	Q9RRZ1	Q9rrz1 deinococcus
14	238	7.7	242	2	Q52660	Q52660 rhodobacter
15	236	7.7	312	16	Q9CNX6	Q9cnx6 pasteurella
16	232.5	7.5	317	16	Q9PCI3	Q9pci3 xylella fas
17	231.5	7.5	248	16	Q9HWC1	Q9hwc1 pseudomonas
18	221	7.2	255	16	Q92QN8	Q92qn8 rhizobium m
19	221	7.2	257	2	Q9R9L1	Q9r9l1 rhizobium m
20	218.5	7.1	274	16	Q9CCL3	Q9ccl3 mycobacteri
21	218.5	7.1	325	16	Q92AA4	Q92aa4 listeria in
22	217.5	7.1	325	16	Q8Y5Z9	Q8y5z9 listeria mo
23	215	7.0	323	16	Q9KC88	Q9kc88 bacillus ha
24	214	6.9	266	16	P96884	P96884 mycobacteri
25	208.5	6.8	274	16	P72816	P72816 synechocyst
26	207	6.7	333	1	Q977S0	Q977s0 uncultured
27	205	6.7	235	17	O73952	O73952 pyrococcus
28	203	6.6	233	16	O66837	O66837 aquifex aeo
29	198.5	6.4	242	16	Q9PCI4	Q9pci4 xylella fas
30	195	6.3	166	5	Q9N802	Q9n802 leishmania
31	193.5	6.3	261	17	O27938	O27938 methanobact
32	193	6.3	291	16	Q9EWV9	Q9ewv9 streptomyce
33	189.5	6.2	237	17	Q9V2C7	Q9v2c7 pyrococcus
34	189	6.1	235	17	O57883	O57883 pyrococcus
35	186.5	6.1	229	1	Q977N1	Q977n1 uncultured
36	183	5.9	273	16	Q97EB4	Q97eb4 clostridium
37	182.5	5.9	323	16	Q99U33	Q99u33 staphylococ
38	181.5	5.9	250	16	Q9A6Z0	Q9a6z0 caulobacter
39	180.5	5.9	265	16	Q9X8N6	Q9x8n6 streptomyce
40	180.5	5.9	325	17	Q8TSW9	Q8tsw9 methanosarc
41	179.5	5.8	326	1	Q9HH16	Q9hh16 methanosarc
42	179	5.8	311	16	Q97NV9	Q97nv9 streptococc
43	177.5	5.8	231	16	Q9WY67	Q9wy67 thermotoga
44	177	5.7	328	16	Q8XML5	Q8xml5 clostridium
45	175.5	5.7	256	16	Q8UFW3	Q8ufw3 agrobacteri

ALIGNMENTS

RESULT 1
Q9JWI7
ID Q9JWI7 PRELIMINARY; PRT; 592 AA.
AC Q9JWI7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BirA bifunctional protein (EC 6.3.4.15).
GN BIRA OR NMA0357.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83659.1; -.
DR HSSP; P06709; 1BIA.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BirA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00121; birA_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Ligase; Complete proteome.
SQ SEQUENCE 592 AA; 64468 MW; 7ED2E0CD3B31C630 CRC64;

```
  Query Match           100.0%;  Score 3080;  DB 16;  Length 592;
  Best Local Similarity  100.0%;  Pred. No. 3.3e-203;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592


RESULT 2
Q9JXF1
ID   Q9JXF1      PRELIMINARY;      PRT;   592 AA.
AC   Q9JXF1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein/Bvg accessory factor.
GN   NMB2075.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002557; AAF42394.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; NMB2075; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  64701 MW;  9E27421DA2B41DE3 CRC64;

  Query Match           98.2%;  Score 3024;  DB 16;  Length 592;
  Best Local Similarity  98.5%;  Pred. No. 2.3e-199;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 3
Q8Y2M3
ID   Q8Y2M3      PRELIMINARY;      PRT;   291 AA.
AC   Q8Y2M3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative bifunctional protein: biotin operon repressor and
DE   biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15).
GN   BIRA OR RSC0312 OR RS03279.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
```

```
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13840.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   291 AA;  30057 MW;  B9F1FD254D8F18C7 CRC64;

  Query Match             10.6%;  Score 326;  DB 16;  Length 291;
  Best Local Similarity   32.9%;  Pred. No. 1.5e-14;
  Matches   95;  Conservative   40;  Mismatches  122;  Indels   32;  Gaps    8;

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELAR------------IAP 105
              ||| |       : :|  |  :|:   :     |:| ::   |              ||
Db         10 WRLTR-------DRIRPTGIAAGWPVEVVESTGSTNADLMAAVRDAAWPVAAGVPAGTAP 62

Qy        106 DKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRR 165
                  : :     |: |||||||  |     |  | ||     |      :|  ||    ||
Db         63 LVGGRVLAALR-QTAGRGRQGRPWDGDHG--LTFSVACAFAAEPAQLAGLSLAVGVAVAE 119

Qy        166 ALSRL----GLKTQ---IKWPNDLVVGRDKLGGILIETVRTG-GKTVAVVGIGINFVLPK 217
              |::      |   |   :|||||: :   || |||:||:|   |:| ||:|||:|   |
Db        120 AVAAYAQTHGGNAQALALKWPNDVQIAGRKLAGILVETMRAAPGQTWAVIGIGLNLERPH 179

Qy        218 EVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDH 277
               :| ||  :||          :   :|  ||: |   | ::   | ||||| : | :
Db        180 ALE-AALGRSLSGVEELVERPEPNAVLAALLSALGEHLQRFGSHGLAPFVAPFAARDAFA 238

Qy        278 GKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDD 326
              |: | | :||  | ||   |:| || | :|:    | : |||:|||: |
Db        239 GRPVRLWQDGAVVLEGVAHGIDAQGRLAIESGGRVQWIHSGEVSLRAAD 287


RESULT 4
Q8X709
ID   Q8X709      PRELIMINARY;      PRT;   321 AA.
AC   Q8X709;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Biotin-[acetylCoA carboxylase] holoenzyme synthetase and biotin operon
DE   repressor (Bifunctional protein BirA biotin-[acetylCoA carboxylase]
DE   holoenzyme synthetase / biotin operon repressor).
GN   BIRA OR Z5544 OR ECS4900.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005629; AAG59173.1; -.
DR   EMBL; AP002567; BAB38323.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   321 AA;  35329 MW;  E1C1BDDD39E2B68D CRC64;

  Query Match             9.9%;  Score 305.5;  DB 16;  Length 321;
  Best Local Similarity   31.0%;  Pred. No. 4.3e-13;
  Matches  103;  Conservative   51;  Mismatches  145;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||      |    :|        ||: |    |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                 |  |: : :|| :  ||:  |    :      |:|  :|:  ||   |:    ||  |
Db         59 -SLPEPIQLLNAEQI--LGQLDGGSVTVLPVIDSTNQYLLD--RIGELKSGDA-CVAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQI 176
                |||||:||||     |  |  |  |  :: |    :| || |  :      | :||  | ::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||      |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |    |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||||  ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 5
Q9ZIU6
ID   Q9ZIU6      PRELIMINARY;      PRT;   288 AA.
AC   Q9ZIU6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative biotin protein ligase.
GN   BIRA.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
```

```
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016461; AAC68833.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase.
SQ   SEQUENCE   288 AA;  29666 MW;  39F6E9CDBC5F6F77 CRC64;

  Query Match             9.6%;  Score 295;  DB 2;  Length 288;
  Best Local Similarity   32.9%;  Pred. No. 1.9e-12;
  Matches   80;  Conservative   43;  Mismatches  106;  Indels   14;  Gaps    5;

Qy         92 SSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYE 151
              |:| ::|  || |   |   :   |||  |||| || | :|:|  ||||  :  | |  :
Db         46 STNADLLARAR-AGGGARPCLLGAHLQETGRGRAGRPWQNRVGAALMFSCAFDVDLPPSQ 104

Qy        152 LGSLSPVAAVACRRALSRL-----GLKTQIKWPNDLVVGRDKLGGILIETVRTGGKT--V 204
              | :|||:| :|   || :|     ||   :|||||:     || |:|:|:||   |
Db        105 LPALSPLAGMAACEALRQLVGQPAGL--LVKWPNDVQWHDAKLAGVLVESVRNPAGTGHA 162

Qy        205 AVVGIGINF----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYAR 260
               |:|:|:|      | : :: : :  :       :|   || |:  :           :
Db        163 VVIGMGLNLHGGDALSQALQRSVADWNDVAGPGEKGPDAAADLVAAVARAWQDAAHTLQQ 222

Qy        261 DGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEI 320
              :||  | | :   :   |: | ::  || :: |   |||  | | :: |:| : |  |||
Db        223 EGFGAFGARFARVDALAGRPVNVIDRGEILYTGAASGVDDHGRLLVQAADGMRPVSVGEI 282

Qy        321 SLR 323
              |:|
Db        283 SIR 285


RESULT 6
Q9KV39
ID   Q9KV39      PRELIMINARY;      PRT;   320 AA.
AC   Q9KV39;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein.
GN   VC0319.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004120; AAF93492.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; VC0319; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   320 AA;  35040 MW;  88E6C496EBEFB5B6 CRC64;

  Query Match             9.3%;  Score 287.5;  DB 16;  Length 320;
  Best Local Similarity   28.6%;  Pred. No. 7.4e-12;
  Matches   97;  Conservative   45;  Mismatches  146;  Indels   51;  Gaps   11;

Qy         11 VLAELADGLPQHVSQ-------LARMADMKPQQLNGFWQQMPAHIRGLLR--------QH 55
              :| :||||   |  :       ::| |  |             ||:|:          |
Db         10 ILKQLADG-DFHSGEVLGAQLGISRAAISK------------HIQGIRDWGVDVFRVQG 55

Qy         56 DGYWRLVRPLAVFDAEGLRELGERSGFQTALKHE------CASSNDEILELARIAPDKAH 109
              || :| : : : |          :|  |: : :         |:|  :|:
Db         56 KGY-QLAQAMTLLD---------QSVIQSQVNNPVELHPIIGSTNQYLLDHVETL---VS 102

Qy        110 KTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSR 169
              |:|:   |: ||||:|| |    |  |  |  |  |          || |  ||  ||
Db        103 GTVCLAEYQASGRGRRGRHWVSPFGANLYLSIYWRLDAGMAAAMGLSLVVGVAIVEALEA 162

Qy        170 LGLK-TQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQS 227
              :|:   ::|||||      || |||:| : : |     |:|:|||  :     |
Db        163 MGVDGVKLKWPNDLYYQDKKLAGILVEMSGQAGAAAHLVIGMGINLAMRDNEGNIDQPWI 222

Qy        228 LFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDG 287
                : :   |   |   |:| ||  | ||   |   ||  :   :   |: | ||  |
Db        223 SLAEVTGQSRIDRNALAINLIAALDRTLRQYEISGMQNFVERWNRWDNFIGRPVKLLM-G 281

Qy        288 ETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDD 326
                |  :|:|  | : ||| || :: : ||||||  :|
Db        282 ANEVRGIERGIDEHGGVLLETEEGLKSFIGGEISLRKND 320


RESULT 7
Q9HWC0
ID   Q9HWC0      PRELIMINARY;      PRT;   312 AA.
AC   Q9HWC0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein.
GN   BIRA OR PA4280.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07668.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   312 AA;  34049 MW;  AC92F3BA43C4BCAF CRC64;

  Query Match             9.1%;  Score 281.5;  DB 16;  Length 312;
  Best Local Similarity   33.1%;  Pred. No. 1.8e-11;
  Matches   90;  Conservative   39;  Mismatches  124;  Indels   19;  Gaps    8;

Qy         57 GYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTH 116
              || ||  ||:: |   :  |     :   :     |:| | :   |:|   :   : |
```

```
Db         52 GY-RLASPLSLLDPRKIDSL-----WPVEIMPTVDSTNSEAMR--RLAAGASTPFVIVAE 103

Qy        117 LQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLK-TQ 175
              |: ||||:|| |    || | :|          ||  || |  :|  |||  ||:: |
Db        104 QQTAGRGRRGRAWVSPYGENLYYSLAITVRGGARELEGLSLVVGLAVLRALESLGVQDTG 163

Qy        176 IKWPNDLVVGRDKLGGILIETVRTGGKTV----AVVGIGINFVLPKEVENAASVQSLFQT 231
              :||||||: |  |: |||:|     |        |:|||||  :    :      || :
Db        164 LKWPNDLLQGGKKIAGILLE---LSGDPADLCHVVIGIGINVNMKTATDIGQPWTSLREA 220

Qy        232 ASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                    | : ||  |  :|   | :::|:|||    |::: |   | | ::|    :
Db        221 LGEL--VDRSRLLSALNRQLADYLDRHSREGFAASREEWESCNLWQG-AQVILSTALSSI 277

Qy        292 EGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              || | ||| :| | ||    :|    ||:|||
Db        278 EGRVLGVDSRGALRLEVDGVEQGFSGGELSLR 309


SULT 8
Y2M4
ID   Q8Y2M4      PRELIMINARY;      PRT;   295 AA.
AC   Q8Y2M4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcription regulation accessory factor transcription
DE   regulator protein.
GN   RSC0311 OR RS03278.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13839.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Complete proteome.
SQ   SEQUENCE   295 AA;  30990 MW;  133074A7764BFCBB CRC64;

  Query Match             9.1%;  Score 279.5;  DB 16;  Length 295;
  Best Local Similarity   29.7%;  Pred. No. 2.3e-11;
  Matches   85;  Conservative   49;  Mismatches  101;  Indels   51;  Gaps   11;

Qy        341 LLLDGGNSRLKWAW----------VENGT-FATVGSAPYRDLSPLGAEWAEKVDGNVRIV 389
              ||:| ||:|:|||||             || :   |: |:  |: |  :| :   |
Db         14 LLIDAGNTRIKWAWTAADVAPPAVAPGGTPWQHAGARPHDQLAELVEDWRDCHAG----A 69

Qy        390 GCA--------VCGEFKKAQVQEQLAR-----KIEWLPSSAQALGIRNHYRHPEEHGSDR 436
              | |        | |   :  :  ::||     ::  : | | | |:|| || | : |:||
Db         70 GMAPPDVWISVVAGPALRDALCARIARVFDGARLRIVASEAAAAGLRNGYRDPAQLGTDR 129

Qy        437 WFNALGSRR-FSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANL- 494
              |  |:|:|  :   | ::|: ||| |:| :  || : || |:|| ||  :|:  || |
Db        130 WVGAVGARHAWPDTALLLVTAGTATTLDIVAPDGRFAGGLILPGLTLMMRALSRNTAQLP 189

Qy        495 ----------NRHAGKRYP-FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVD 542
                       :  |    | :   | :|:| | : |  |::      |: :     |
Db        190 EIDIGYLAARDDAQAPADVPSWADNTQDAIALGCVTAQAGAIAQTWQALQAQYPG--PYR 247

Qy        543 VIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGE 588
               :::||  | :|  |   |      :: ||||: | | ::|  | |
Db        248 CVLSGGARAALAPHLRMPF------QMHDNLVLLG-LQVLAHTGHE 286


RESULT 9
Q8ZAN5
ID   Q8ZAN5      PRELIMINARY;      PRT;   319 AA.
AC   Q8ZAN5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional protein BirA (EC 6.3.4.15).
GN   BIRA OR YPO3759.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414158; CAC93227.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   319 AA;  35164 MW;  B58E9306939E3B2A CRC64;

  Query Match             9.0%;  Score 276.5;  DB 16;  Length 319;
  Best Local Similarity   29.0%;  Pred. No. 4.2e-11;
  Matches   96;  Conservative   46;  Mismatches  150;  Indels   39;  Gaps   12;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFD 69
              |::: |:||      ||     |   :|  |     ||         |     | ||
Db          9 RLVSILSDGFFHSGEQLGETLGMSRAAINKHMQT----IRD--------W----GLDVFT 52

Qy         70 AEG--------LRELGER-------SGFQTALKHECASSNDEILELARIAPDKAHKTICV 114
                |         :: | |:       :|  | |     |:|  :|:  ||   |:   ||
Db         53 VPGKGYSLPTSIQLLDEQKILSYLPTGHVTVLP-VVDSTNQYLLD--RITELKSGDA-CV 108

Qy        115 THLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-K 173
              |  ||||:||:|    |  |  |  |  ::       || |  :    |  ||  |
Db        109 AEYQQAGRGRRGRQWVSPFGANLYLSMYWRLEQGPAAAMGLSLVVGIVMAEVLQNLGAEK 168

Qy        174 TQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTA 232
               ::|||||| :   || |||:| | :||     |:| |||  :     || : |
Db        169 VRVKWPNDLYLNDKKLAGILVELTGKTGDAAQLVIGAGINLTMRGSETNAIN-QDWINLQ 227

Qy        233 SRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFE 292
                    |   |   :|:||   ::::  :| : |:: :|  :    : | |:   : :|
Db        228 DAGVTIDRNKLTAEILSELRLAVVKFENEGLSAFLSRWQKMDNYLNRPVKLIIGNQEIF- 286

Qy        293 GTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              |  :|:| || | ||     :  : ||||||
Db        287 GIARGIDQQGALLLEQDGNIKPYIGGEISLR 317


RESULT 10
Q8Z317
ID   Q8Z317      PRELIMINARY;      PRT;   320 AA.
```

```
AC   Q8Z317;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional protein: biotin operon repressor and biotin-[acetyl-CoA
DE   carboxylase] synthetase.
GN   STY3741.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627279; CAD09497.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   320 AA;  35396 MW;  B584A48BABBEF59F CRC64;

  Query Match             8.8%;  Score 269.5;  DB 16;  Length 320;
  Best Local Similarity   29.6%;  Pred. No. 1.3e-10;
  Matches   97;  Conservative   50;  Mismatches  148;  Indels   33;  Gaps   13;

Qy         11 VLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GYW 59
              ::: ||||      ||     |     :|        ||: |    |          ||
Db         10 LISLLADGEFHSGEQLGERLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY- 58

Qy         60 RLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQS 119
               |  |: : ||: :    :      |:     |:|  :|:  ||   ::    ||   |
Db         59 SLPEPIQLLDADRIH--SQLDSGNVAVLPVIDSTNQYLLD--RIGELRSGDA-CVAEYQQ 113

Qy        120 KGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQIK 177
              ||||:||||    |  |  |  |  :: |   :| || |  :     |  ||  | ::|
Db        114 AGRGRRGRKWFSPFGANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRELGADKVRVK 172

Qy        178 WPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRG 236
              ||||| :   || |||:| | :||     |:| |||  : : ||      |
Db        173 WPNDLYLLDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGI 231

Qy        237 NADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVK 296
               |  :|   |: :| | |  : ::| :|::: ::  :    : | |:  : :| |  :
Db        232 TLDRNMLAAKLIYKLRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLIIGDKEIF-GISR 290

Qy        297 GVDGQGVLHLETAEGKQTVVSGEISLRS 324
              |:| || | ||     :  : |||||||
Db        291 GIDTQGALLLEQDGVIKPWMGGEISLRS 318


RESULT 11
Q8R8N0
ID   Q8R8N0      PRELIMINARY;      PRT;   326 AA.
AC   Q8R8N0;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Biotin-(acetyl-CoA carboxylase) ligase.
GN   BIRA OR TTE1965.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013148; AAM25144.1; -.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   326 AA;  37019 MW;  CD98EF8D22966E2B CRC64;

  Query Match             8.0%;  Score 246;  DB 16;  Length 326;
  Best Local Similarity   32.3%;  Pred. No. 5.4e-09;
  Matches   81;  Conservative   43;  Mismatches   97;  Indels   30;  Gaps   13;

Qy         87 KHECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFD 146
              |   :|:||   |||   ||    |: |   |:|||| || |       |:: |   |
Db         86 KESLSSTNDMAKELAYKVPD---GTVIVAEEQTKGRGRMGRSWFSEKSSCILTS---VIL 139

Qy        147 RPQY---ELGSLSPVAAVACRRALSRL-GLKTQIKWPNDLVVGRDKLGGILIE-TVRTGG 201
              :||    :  :|: |||:   :|:  :  :||:||||||:::   |: ||| | :
Db        140 KPQIRPEKTVALTQVAALTVVKAIEEVCSVKTKIKWPNDIILNSKKVCGILTEMSSEIDV 199

Qy        202 KTVAVVGIGINF---VLPKEV-ENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 257
                |:|||||      | :: | | |:|   :        |   || :||  |:    :
Db        200 VNYIVIGIGINVNCTEFPDDLKEKATSLQLELEKP-----VDRKKLLASLLNNLEIYYKE 254

Qy        258 YARDGF---APFVAEYQAANRDHGKAV-LLLRDGETVFEGTVKGVDGQGVLHLETAEGKQ 313
              |   ||    | : |          || | ::  | |     |    :| :| | ::| :|::
Db        255 YEEKGFESLRPLILENSIT---IGKEVRVIYTDRE--IRGQAIAIDKEGKLVIKTEKGEE 309

Qy        314 -TVVSGEISLR 323
               ::|||:|:|
Db        310 IALLSGEVSVR 320


RESULT 12
O32511
ID   O32511      PRELIMINARY;      PRT;   322 AA.
AC   O32511;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   ORF105d (Fragment).
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KD8301;
RX   MEDLINE=20075120; PubMed=10606814;
RA   Narumi I., Satoh K., Kikuchi M., Funayama T., Kitayama S.,
RA   Yanagisawa T., Watanabe H., Yamamoto K.;
RT   "Molecular analysis of the Deinococcus radiodurans recA locus and
RT   identification of a mutation site in a DNA repair-deficient mutant,
RT   rec30.";
RL   Mutat. Res. 435:233-243(1999).
DR   EMBL; AB005471; BAA21331.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
```

```
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
FT   NON_TER       1       1
SQ   SEQUENCE   322 AA;  33999 MW;  183BCDA5CBA9C31B CRC64;

  Query Match             7.9%;  Score 244;  DB 2;  Length 322;
  Best Local Similarity   32.5%;  Pred. No. 7.3e-09;
  Matches  111;  Conservative   41;  Mismatches  122;  Indels   68;  Gaps   20;

Qy          9 WRVLAELADGL---PQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPL 65
              ||: | |    |   ||    | |  |:    :|   :::          | ||      |:
Db         15 WRMPARLLPLLTDRPQSGDDLGRALDLGRVTVNTLARRL---------QEDGV-----PV 60

Qy         66 AVFDAEGLRELGE--------RSGFQTALKHE--CASSNDEILELARIAPDK-----AHK 110
               :  |    | |          |  |  ||:::    |: ||:    |   |        |
Db         61 LISRAGYALEPGTPAPGLVPLRGQFGRALRYQGTVTSTQDEV----RFWADSERGPAPHG 116

Qy        111 TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSL--SPVAA-VACRRAL 167
               : |    |: ||||:|| |    |  | ||   |  |     |  |:|| ||   |
Db        117 AVVVAERQTAGRGRRGRVWDTTHG-TLAFS---VLLREALSLADLVRMPLAAGVALHAAC 172

Qy        168 SRLGLKTQIKWPNDLVV--GRDKLGGILIETVRTGGKT-VAVVGIGINFVLPKEVENAAS 224
              : ||:   :||||||:   || || |||:|    | :   ||:|||||  :    | ||
Db        173 TALGVPCGLKWPNDLLTPDGR-KLAGILLEADLRGEEARRAVLGIGIN--VGTAPEGAAH 229

Qy        225 VQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLL 284
              :       :|     |||| | |  :|:  | | |       :| ::| :   |: | :
Db        230 LHESVPDLTR-----AAVLGEVLF-QLEHWLAQPA----PAIIATWKATSLTLGRPVRV- 278

Qy        285 RDGETVFEGTVKGV----DGQGVLHLETAEGKQTVVSGEISL 322
                  |   |||:||    | || | :||| | :|| :|:: |
Db        279 ----TTARGTVEGVARDLDEQGNLLVETAAGLETVSAGDVQL 316


RESULT 13
Q9RRZ1
ID   Q9RRZ1      PRELIMINARY;      PRT;   474 AA.
AC   Q9RRZ1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BIRA bifunctional protein.
GN   DR2341.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
OC   Deinococcaceae; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
DR   EMBL; AE002065; AAF11888.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; DR2341; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   474 AA;  49966 MW;  58E1E537E0D9B0EE CRC64;

  Query Match             7.9%;  Score 244;  DB 16;  Length 474;
  Best Local Similarity   32.5%;  Pred. No. 1.3e-08;
  Matches  111;  Conservative   41;  Mismatches  122;  Indels   68;  Gaps   20;

Qy          9 WRVLAELADGL---PQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPL 65
              ||: | |    |   ||    | |  |:    :|   :::          | ||      |:
Db        167 WRMPARLLPLLTDRPQSGDDLGRALDLGRVTVNTLARRL---------QEDGV-----PV 212

Qy         66 AVFDAEGLRELGE--------RSGFQTALKHE--CASSNDEILELARIAPDK-----AHK 110
               :  |    | |          |  |  ||:::    |: ||:    |   |        |
Db        213 LISRAGYALEPGTPAPGLVPLRGQFGRALRYQGTVTSTQDEV----RFWADSERGPAPHG 268

Qy        111 TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSL--SPVAA-VACRRAL 167
               : |    |: ||||:|| |    |  | ||   |  |     |  |:|| ||   |
Db        269 AVVVAERQTAGRGRRGRVWDTTHG-TLAFS---VLLREALSLADLVRMPLAAGVALHAAC 324

Qy        168 SRLGLKTQIKWPNDLVV--GRDKLGGILIETVRTGGKT-VAVVGIGINFVLPKEVENAAS 224
              : ||:   :||||||:   || || |||:|    | :   ||:|||||  :    | ||
Db        325 TALGVPCGLKWPNDLLTPDGR-KLAGILLEADLRGEEARRAVLGIGIN--VGTAPEGAAH 381

Qy        225 VQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLL 284
              :       :|     |||| | |  :|:  | | |       :| ::| :   |: | :
Db        382 LHESVPDLTR-----AAVLGEVLF-QLEHWLAQPA----PAIIATWKATSLTLGRPVRV- 430

Qy        285 RDGETVFEGTVKGV----DGQGVLHLETAEGKQTVVSGEISL 322
                  |   |||:||    | || | :||| | :|| :|:: |
Db        431 ----TTARGTVEGVARDLDEQGNLLVETAAGLETVSAGDVQL 468


RESULT 14
Q52660
ID   Q52660      PRELIMINARY;      PRT;   242 AA.
AC   Q52660;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BIRA.
GN   BIRA.
OS   Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Rhodobacter.
OX   NCBI_TaxID=1061;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B10;
RX   MEDLINE=96144256; PubMed=8566820;
RA   Dupuis A., Peinnequin A., Chevallet M., Lunardi J., Darrouzet E.,
RA   Pierrard B., Procaccio V., Issartel J.P.;
RT   "Identification of five Rhodobacter capsulatus genes encoding the
RT   equivalent of ND subunits of the mitochondrial NADH-ubiquinone
RT   oxidoreductase.";
RL   Gene 167:99-104(1995).
DR   EMBL; AF029365; AAC25006.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
SQ   SEQUENCE   242 AA;  25117 MW;  6484B805B7DDE080 CRC64;

  Query Match             7.7%;  Score 238;  DB 2;  Length 242;
  Best Local Similarity   32.0%;  Pred. No. 1.2e-08;
  Matches   80;  Conservative   34;  Mismatches  108;  Indels   28;  Gaps    9;

Qy         87 KHECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFD 146
              :|    ::  : | ||:||  |       | |: ||||:|| |    |    |:   ||
Db          3 RHVLGQTDTTMAEAARLAPHLAGPAWVFAHRQTAGRGRRGRDWLMPSGN---FAATLVF- 58
```

```
Qy        147 RPQYELGSLSP--------VAAVACRRALSRL---GLKTQIKWPNDLVVGRDKLGGILIE 195
              ||     :::|        |||:|   ||: :   |  | ||||||:::   |: |||:|
Db         59 RP-----AVTPATAALYSFVAALALEAALTEVMGPGPVTAIKWPNDVLLNGGKIAGILLE 113

Qy        196 TVRTGGK-TVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAV 254
              :| |||: :   ||||:|  | |    ||          |  |     :     || |
Db        114 SVGTGGQISHLAVGIGVN--LAKAPPAAALEPGAVAPVSLLGETGLVIDPLDFLAPLARG 171

Qy        255 LLQYA----RDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAE 310
              ||    | ||||  | : |     |: :   | |     || : :|  | | | |
Db        172 FAHYAATFERYGFAPIRAAWLARAARIGQRI-TARTGTETLTGTFETIDDTGALILATGS 230

Qy        311 GKQTVVSGEI 320
              :: : : :|
Db        231 ERRAISAADI 240


RESULT 15
Q9CNX6
    Q9CNX6      PRELIMINARY;      PRT;  312 AA.
    Q9CNX6;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT  01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE  BirA.
GN  BIRA OR PM0296.
OS  Pasteurella multocida.
OC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC  Pasteurella.
OX  NCBI_TaxID=747;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=PM70;
RX  MEDLINE=21145866; PubMed=11248100;
RA  May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT  "Complete genomic sequence of Pasteurella multocida Pm70.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR  EMBL; AE006065; AAK02380.1; -.
DR  HSSP; P06709; 1BIA.
DR  InterPro; IPR004408; BirA_ligase.
DR  InterPro; IPR003142; BPL_C.
DR  InterPro; IPR004143; BPL_LipA_LipB.
DR  Pfam; PF02237; BPL_C; 1.
DR  Pfam; PF03099; BPL_LipA_LipB; 1.
DR  TIGRFAMs; TIGR00121; birA_ligase; 1.
KW  Complete proteome.
SQ  SEQUENCE   312 AA;  35214 MW;  0DD8B737C7C0E29B CRC64;

  Query Match           7.7%;  Score 236;  DB 16;  Length 312;
  Best Local Similarity   29.1%;  Pred. No. 2.5e-08;
  Matches   84;  Conservative   47;  Mismatches  118;  Indels   40;  Gaps   11;

Qy         52 LRQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAH-- 109
              : |   ::||:  |:  :   | :    : :   :|    |:|  :|:      : ||
Db         44 IEQQVQHFRLIPQLSRLNESYLTQ--ALAPYPLYIKPIINSTNQFLLD------NIAHLE 95

Qy        110 -KTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALS 168
                ||:   |: ||||:||:|       :: |  |  |: :  |  || |  :|    |
Db         96 KGAICLAEYQTAGRGRRGRQWLSPFAGQVIMSLYWTCDQ-KVNLEGLSLVVGMAIAETLK 154

Qy        169 RLG-LKTQIKWPNDLVVGRDKLGGILIETV-RTGGKTVAVVGIGINFVLPKEVENA---- 222
              : | |   :|||||:::   || |||:|   |   :   |:| |||   ||: :
Db        155 QAGALNIGLKWPNDVLLHGRKLAGILVEIANRQNNQHNLVIGFGINLSFPKQTQQIYQPW 214

Qy        223 ASVQSLFQTASRRGNADAAVLLETLLAEL--------DAVLLQYARDGFAPFVAEYQAAN 274
              | :   :  |  |  |   | | : |:| |       |    |  ||  | |
Db        215 AELIEVLPTIDR--NKLIAELGKKLIARLQHFEQAGIDEAFRQDWRDMDAYF-------- 264

Qy        275 RDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
                |: | :| : : :  |  :|:| :| | | |:|| |    ||:|||
Db        265 ---GEEVHVLTENQKI-TGIAQGIDKRGYLQLITSEGIQYFNGGEVSLR 309
```

Search completed: June 24, 2003, 21:59:39
Job time : 50.2837 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 10.2965 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-22
Perfect score: 3080
Sequence: 1 MTVLKPSHWRVLAELADGLP.....LVIHGLLNLIAAEGGESEHT 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

otal number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	304.5	9.9	321	1	BIRA_ECOLI	P06709 escherichia
2	271.5	8.8	320	1	BIRA_SALTY	P37416 salmonella
3	256.5	8.3	267	1	BAF_BORPE	Q45338 bordetella
4	256	8.3	302	1	BIRA_HAEIN	P46363 haemophilus
5	226.5	7.4	325	1	BIRA_BACSU	P42975 bacillus su
6	197.5	6.4	240	1	BIRA_PARDE	P29906 paracoccus
7	133	4.3	233	1	YACB_BACSU	P37564 bacillus su
8	130.5	4.2	1655	1	OMPB_RICCN	Q9kka3 r outer mem
9	122	4.0	1132	1	VHSJ_LAMBD	P03749 bacteriopha
10	121.5	3.9	726	1	BPL1_HUMAN	P50747 h biotin--p
11	121	3.9	471	1	GLTD_ECOLI	P09832 escherichia
12	118	3.8	1656	1	OMPB_RICJA	O06653 r outer mem
13	112.5	3.7	1654	1	OMPB_RICRI	Q53047 r outer mem
14	110.5	3.6	776	1	HYPF_AZOCH	Q43950 azotobacter
15	110	3.6	356	1	FRVX_ECOLI	P32153 escherichia
16	109.5	3.6	806	1	HYF2_ALCEU	O07451 alcaligenes
17	109.5	3.6	1609	1	CTPI_MYCLE	O53114 mycobacteri
18	109	3.5	752	1	CTPB_MYCTU	Q10877 mycobacteri
19	108.5	3.5	573	1	CH60_DROME	O02649 drosophila
20	106.5	3.5	1756	1	TRI2_ECOLI	P22706 escherichia
21	105.5	3.4	691	1	YHJG_ECOLI	P37645 escherichia
22	105	3.4	580	1	RUBA_CHLRE	Q42694 chlamydomon
23	105	3.4	721	1	ATZN_SYNY3	Q59998 synechocyst
24	105	3.4	1524	1	RPOC_THEAQ	Q9kwu6 thermus aqu
25	104.5	3.4	3172	1	ERY3_SACER	Q03133 saccharopol
26	103.5	3.4	548	1	FUMB_ECOLI	P14407 escherichia
27	103.5	3.4	700	1	TKT_MYCTU	O06811 mycobacteri
28	102.5	3.3	690	1	BPL1_YEAST	P48445 s biotin--p
29	102.5	3.3	1127	1	LONH_PYRHO	O58221 pyrococcus
30	102	3.3	775	1	PLSB_MYCLE	Q9x7b0 mycobacteri
31	101.5	3.3	506	1	ATPA_ANASP	P12405 anabaena sp
32	101.5	3.3	1538	1	PPSB_MYCTU	Q10978 mycobacteri
33	101	3.3	906	1	HELY_MYCTU	Q10701 mycobacteri
34	100.5	3.3	417	1	SAHH_SULSO	P50252 sulfolobus
35	100.5	3.3	544	1	CH60_NEIFL	P48215 neisseria f
36	100.5	3.3	739	1	HYPF_RHOCA	Q02987 rhodobacter
37	100.5	3.3	3674	1	SPCR_HUMAN	Q9nrc6 homo sapien
38	99.5	3.2	544	1	CH60_NEIMA	P57006 neisseria m
39	98.5	3.2	841	1	RELA_STRAT	O85709 streptomyce
40	98.5	3.2	1059	1	Y379_HUMAN	O15084 homo sapien
41	98	3.2	477	1	DLDH_AZOVI	P18925 azotobacter
42	98	3.2	1448	1	TRC5_ECOLI	P27190 escherichia
43	97.5	3.2	300	1	RNH3_STRPY	Q99y70 streptococc
44	97.5	3.2	1316	1	RPOC_MYCTU	P47769 mycobacteri
45	97	3.1	508	1	GLPK_MYCLE	Q9cb81 mycobacteri

ALIGNMENTS

RESULT 1
BIRA_ECOLI
ID BIRA_ECOLI STANDARD; PRT; 321 AA.
AC P06709;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE BirA bifunctional protein [Includes: Biotin operon repressor;
DE Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE (Biotin--protein ligase)].
GN BIRA OR BIOR OR DHBB OR B3973.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86006282; PubMed=3899863;
RA Howard P.K., Shaw J., Otsuka A.J.;
RT "Nucleotide sequence of the birA gene encoding the biotin operon
RT repressor and biotin holoenzyme synthetase functions of Escherichia
RT coli.";
RL Gene 35:321-331(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87055242; PubMed=3536662;
RA Buoncristiani M.R., Howard P.K., Otsuka A.J.;
RT "DNA-binding and enzymatic domains of the bifunctional biotin operon
RT repressor (BirA) of Escherichia coli.";
RL Gene 44:255-261(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94089392; PubMed=8265357;
RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA Daniels D.L.;
RT "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT region from 89.2 to 92.8 minutes.";
RL Nucleic Acids Res. 21:5408-5417(1993).
RN [4]
RP SEQUENCE OF 1-76 FROM N.A.
RC STRAIN=RDD012;
RX MEDLINE=92165747; PubMed=1311302;
RA Pucci M.J., Discotto L.F., Dougherty T.J.;
RT "Cloning and identification of the Escherichia coli murB DNA
RT sequence, which encodes UDP-N-acetylenolpyruvoylglucosamine
RT reductase.";
RL J. Bacteriol. 174:1690-1693(1992).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=93028443; PubMed=1409631;
RA Wilson K.P., Shewchuk L.M., Brennan R.G., Otsuka A.J., Matthews B.W.;
RT "Escherichia coli biotin holoenzyme synthetase/bio repressor crystal
RT structure delineates the biotin- and DNA-binding domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9257-9261(1992).
CC -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS

```
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10123; AAA23520.1; -.
DR   EMBL; M15820; AAA23521.1; -.
DR   EMBL; L14557; AAA24186.1; -.
DR   EMBL; U00006; AAC43075.1; -.
DR   EMBL; AE000471; AAC76951.1; -.
DR   PIR; B24029; BVECBF.
DR   PIR; A24989; A24989.
DR   PDB; 1BIA; 31-OCT-93.
DR   PDB; 1BIB; 31-OCT-93.
DR   EcoGene; EG10123; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   3D-structure; Complete proteome.
FT   DNA_BIND      22     41       H-T-H MOTIF.
FT   VARIANT       61     61       P -> A (IN STRAIN RDD012).
FT   VARIANT       70     70       K -> E (IN STRAIN RDD012).
FT   HELIX          5     14
FT   TURN          15     16
FT   STRAND        20     20
FT   HELIX         22     29
FT   TURN          30     30
FT   HELIX         33     45
FT   TURN          46     47
FT   STRAND        51     53
FT   TURN          54     56
FT   STRAND        57     59
FT   HELIX         69     74
FT   TURN          75     75
FT   STRAND        81     83
FT   STRAND        86     87
FT   HELIX         90    100
FT   TURN         103    104
FT   STRAND       106    109
FT   STRAND       111    112
FT   TURN         127    128
FT   STRAND       130    139
FT   TURN         146    146
FT   HELIX        147    163
FT   TURN         164    165
FT   TURN         167    168
FT   STRAND       170    172
FT   TURN         173    175
FT   STRAND       176    179
FT   TURN         180    181
FT   STRAND       182    192
FT   STRAND       199    208
FT   TURN         226    226
FT   HELIX        227    229
FT   TURN         230    230
FT   HELIX        235    254
FT   TURN         255    256
FT   HELIX        259    268
FT   STRAND       269    269
FT   TURN         272    273
FT   STRAND       275    280
FT   TURN         281    282
FT   STRAND       283    292
FT   TURN         294    295
FT   STRAND       298    302
FT   TURN         303    304
FT   STRAND       305    309
FT   STRAND       313    313
FT   STRAND       316    316
SQ   SEQUENCE   321 AA;  35312 MW;  B80AEBCEEE1BD2D4 CRC64;

  Query Match             9.9%;  Score 304.5;  DB 1;  Length 321;
  Best Local Similarity   30.7%;  Pred. No. 1e-14;
  Matches  102;  Conservative   53;  Mismatches  144;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||      |    :|        ||: |    |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|: :  ||:  |   |:     |:|  :|:  ||   |:    |:   |
Db         59 -SLPEPIQLLNAKQI--LGQLDGGSVAVLPVIDSTNQYLLD--RIGELKSGDA-CIAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQI 176
                ||||:||||    |  |  |  :: |   :| || |  :     | :||  | ::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
              | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : ||||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 2
BIRA_SALTY
ID   BIRA_SALTY     STANDARD;      PRT;   320 AA.
AC   P37416; Q9L9K4;
DT   01-OCT-1994 (Rel. 30, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR STM4138 OR STMF1.2.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
```

```
RL   Nature 413:852-856(2001).
RN   [2]
RP   SEQUENCE OF 1-135 FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=94324848; PubMed=8048842;
RA   Dombrosky P.M., Schmid M.B., Young K.D.;
RT   "Sequence divergence of the murB and rrfB genes from Escherichia coli
RT   and Salmonella typhimurium.";
RL   Arch. Microbiol. 161:501-507(1994).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF170176; AAF33493.1; -.
DR   EMBL; AE008893; AAL22971.1; -.
DR   EMBL; L14816; -; NOT_ANNOTATED_CDS.
DR   HSSP; P06709; 1BIA.
DR   StyGene; SG10028; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
FT   DNA_BIND     22     41       H-T-H MOTIF.
FT   CONFLICT    100    101       EL -> DV (IN REF. 2).
SQ   SEQUENCE   320 AA;  35395 MW;  010E8483093297F5 CRC64;

  Query Match             8.8%;  Score 271.5;  DB 1;  Length 320;
  Best Local Similarity   29.6%;  Pred. No. 2.4e-12;
  Matches   97;  Conservative   51;  Mismatches  147;  Indels   33;  Gaps   13;

Qy         11 VLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GYW 59
              ::: ||||      ||      |    :|        ||: |    |          ||
Db         10 LISLLADGEFHSGEQLGERLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY- 58

Qy         60 RLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQS 119
                |  |: : ||: :    :      |:     |:|  :|:  ||   ::    ||  |
Db         59 SLPEPIQLLDADRIH--SQLDSGNVAVLPVIDSTNQYLLD--RIGELRSGDA-CVAEYQQ 113

Qy        120 KGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQIK 177
               ||||:||||    |  |  |  |. :: |   :| || |  :     | :||  | ::|
Db        114 AGRGRRGRKWFSPFGANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVK 172

Qy        178 WPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRG 236
               |||||  :   || |||:| | :||     |:| |||  : : ||     |
Db        173 WPNDLYLLDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGI 231

Qy        237 NADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVK 296
                |  :|   |: :| | |  : ::| :|::: ::  :    : | |:   : :|.|  :
Db        232 TLDRNMLAAKLIYKLRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLIIGDKEIF-GISR 290

Qy        297 GVDGQGVLHLETAEGKQTVVSGEISLRS 324
```

```
                |:| || | ||     :  : |||||||
Db        291 GIDTQGALLLEQDGVIKPWMGGEISLRS 318


RESULT 3
BAF_BORPE
ID   BAF_BORPE      STANDARD;      PRT;   267 AA.
AC   Q45338; Q45373;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Bvg accessory factor.
GN   BAF.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RX   MEDLINE=95325323; PubMed=7601846;
RA   Deshazer D., Wood G.E., Friedman R.L.;
RT   "Identification of a Bordetella pertussis regulatory factor required
RT   for transcription of the pertussis toxin operon in Escherichia
RT   coli.";
RL   J. Bacteriol. 177:3801-3807(1995).
RN   [2]
RP   SEQUENCE OF 1-38 FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 239-267 FROM N.A.
RC   STRAIN=BP536;
RX   MEDLINE=96419162; PubMed=8821935;
RA   Allen A.G., Maskell D.J.;
RT   "The identification, cloning and mutagenesis of a genetic locus
RT   required for lipopolysaccharide biosynthesis in Bordetella
RT   pertussis.";
RL   Mol. Microbiol. 19:37-52(1996).
CC   -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC       A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC       RNA POLYMERASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U12020; AAA75361.1; -.
DR   EMBL; AF016461; AAC68834.1; -.
DR   EMBL; X90711; CAA62242.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Transcription regulation; Activator.
SQ   SEQUENCE   267 AA;  27845 MW;  336A615F67B57901 CRC64;

  Query Match             8.3%;  Score 256.5;  DB 1;  Length 267;
  Best Local Similarity   32.0%;  Pred. No. 2.2e-11;
  Matches   85;  Conservative   43;  Mismatches  113;  Indels   25;  Gaps    9;

Qy        341 LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKVDGNVRIVGCAVCG 395
              :|:| ||||||  | : :   |   ||      ||  ||   |     | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        396 EFKKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 449
                :   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122
```

```
Qy        450 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTG 509
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        510 NAVASGMMDAVCGSVM--MMHGRLKEKTGAGKPVDVIITGGG-------AAKVAEALPPA 560
Db        183 QAIASGIAAAQAGAIVRQWLAGRQR----YGQAPEIYVAGGGWPEVRQEAERLLAVTGAA 238

Qy        561 FLAENTVRVADNLVIHGLLNLIAAEG 586
Db        239 FGATPQPTYLDSPVLDGLAAL-AAQG 263


RESULT 4
BIRA_HAEIN
ID   BIRA_HAEIN     STANDARD;      PRT;   302 AA.
AC   P46363;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR HI0220.1.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION.
RA   Koonin E.V., Rudd K.E.;
RL   Submitted (SEP-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32708; AAC21889.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; HI0220.1; -.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
SQ   SEQUENCE   302 AA;  34341 MW;  50EE9D6E9820378B CRC64;

  Query Match             8.3%;  Score 256;  DB 1;  Length 302;
  Best Local Similarity   29.1%;  Pred. No. 2.8e-11;
  Matches   91;  Conservative   56;  Mismatches  132;  Indels   34;  Gaps   13;

Qy         26 LARMADMKPQ---QLNGFWQQMPAHIR-----GLLRQHDGY-WRLVRPLAVFDAEGLREL 76
Db          7 LTYLADCQPKVRSELEKFSKNLEEDIQQLREIGLDILVDGQDYRLVPMLPLLNPQQI--- 63

Qy         77 GERSGFQTALKHE--CASSNDEILE--LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHR 132
Db         64 -STALFPYSIHYQPIISSTNEWILQNILSLKKGD-----LCVAEYQTAGRGRRGRQWLSP 117

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192
Db        118 FAGQIMFSFYWAFD-PKKSIEGLSLVIGLAIAEVLN-----VQVKWPNDILFDERKLGGI 171

Qy        193 LIETV-RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 251
Db        172 LVEIANHKNGMLNLVIGIGINVSLSKQTEIS---QPYAEVCEIDPDVERQTLLPKLIQHL 228

Qy        252 DAVLLQYARDGF-APFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAE 310
Db        229 YTRLNIFEQNGIDEEFQQAWQSYNAFSNSEINVLTE-QGVISGIEQGIDERGYLKVLCGN 287

Qy        311 GKQTVVSGEISLR 323
Db        288 KIQMFNGGEVSLR 300


RESULT 5
BIRA_BACSU
ID   BIRA_BACSU     STANDARD;      PRT;   325 AA.
AC   P42975;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg, and W23 / PY79;
RX   MEDLINE=95247697; PubMed=7730294;
RA   Bower S., Perkins J.P., Yocum R.R., Serror P., Sorokin A.V.,
RA   Rahaim P., Howitt C.L., Prasad N., Ehrlich S.D., Pero J.;
RT   "Cloning and characterization of the Bacillus subtilis birA gene
RT   encoding a repressor of the biotin operon.";
RL   J. Bacteriol. 177:2572-2575(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=96349105; PubMed=8760912;
RA   Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA   Serror P.;
RT   "Sequence analysis of the Bacillus subtilis chromosome region between
RT   the serA and kdg loci cloned in a yeast artificial chromosome.";
RL   Microbiology 142:2005-2016(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38424; AAA92879.1; ALT_SEQ.
DR   EMBL; U20445; AAB60184.1; -.
DR   EMBL; L47709; AAB38447.1; -.
DR   EMBL; Z99115; CAB14160.1; -.
DR   HSSP; P06709; 1BIB.
DR   SubtiList; BG11206; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
FT   DNA_BIND      23     42       H-T-H MOTIF (BY SIMILARITY).
FT   VARIANT       18     18       N -> S (IN STRAIN W23).
FT   VARIANT       89     89       H -> Q (IN STRAIN W23).
FT   VARIANT       92     92       L -> I (IN STRAIN W23).
FT   VARIANT      137    137       V -> I (IN STRAIN W23).
FT   VARIANT      169    169       E -> A (IN STRAIN W23).
FT   VARIANT      219    220       PN -> SD (IN STRAIN W23).
FT   VARIANT      237    237       A -> S (IN STRAIN W23).
FT   VARIANT      311    311       N -> Q (IN STRAIN W23).
FT   MUTAGEN       23     23       G->S: DEREGULATION OF BIOTIN SYNTHESIS.
FT   MUTAGEN       38     38       W->R: DEREGULATION OF BIOTIN SYNTHESIS.
FT   MUTAGEN       58     58       G->E: DEREGULATION OF BIOTIN SYNTHESIS.
SQ   SEQUENCE   325 AA;  36182 MW;  69088BB2CAD09221 CRC64;

  Query Match             7.4%;  Score 226.5;  DB 1;  Length 325;
  Best Local Similarity   28.0%;  Pred. No. 4e-09;
  Matches   86;  Conservative   50;  Mismatches  118;  Indels   53;  Gaps   13;

Qy         47 HIRGLLRQ--------HDGYWRLVRPLAVFDAE---GLRE--LGERSGFQTALK------ 87
              ||  | ::          ||  : :|  : ::|   ||:   :|:   :   |
Db         40 HIEELRKEGYEVEAVRRKGYRLIKKPGKLSESEIRFGLKTEVMGQHLIYHDVLSSTQKTA 99

Qy         88 HECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDR 147
              || |::|          ||:    |: |   |: |||| | |  : |  :  |.    |
Db        100 HELANNN---------APE---GTLVVADKQTAGRGRMSRVWHSQEGNGVWMSLILRPDI 147

Qy        148 PQYELGSLSPVAAVACRRALSR-LGLKTQIKWPNDLVVGRDKLGGILIE------TVRTG 200
              |  :   |: :|||| : :    |::| ||||||:::   |  ||| |       ||:
Db        148 PLQKTPQLTLLAAVAVVQGIEEAAGIQTDIKWPNDILINGKKTVGILTEMQAEEDRVRS- 206

Qy        201 GKTVAVVGIGINF-----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVL 255
                ::|||||        | |::: |:  || |,|  :   | | ::: :|   :
Db        207 ----VIIGIGINVNQQPNDFPDELKDIAT--SLSQAAGEK--IDRAGVIQHILLCFEKRY 258

Qy        256 LQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTV 315
              |   || |    :::    |   :  |    | |   |:| :||| ||| || : :
Db        259 RDYMTHGFTPIKLLWESYALGIG-TNMRARTLNGTFYGKALGIDDEGVLLLETNEGIKKI 317

Qy        316 VSGEISL 322
               | :| |
Db        318 YSADIEL 324


RESULT 6
BIRA_PARDE
ID   BIRA_PARDE     STANDARD;      PRT;   240 AA.
AC   P29906;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15) (Biotin--
DE   protein ligase).
GN   BIRA.
OS   Paracoccus denitrificans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Paracoccus.
OX   NCBI_TaxID=266;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13543;
RX   MEDLINE=93136200; PubMed=8422400;
RA   Xu X., Matsuno-Yagi A., Yagi T.;
RT   "DNA sequencing of the seven remaining structural genes of the gene
RT   cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT   of Paracoccus denitrificans.";
RL   Biochemistry 32:968-981(1993).
CC   -!- FUNCTION: ACTIVATES BIOTIN TO FORM BIOTINYL-5'-ADENYLATE AND
CC       TRANSFERS THE BIOTIN MOIETY TO BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L02354; AAA25601.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Biotin; Ligase.
SQ   SEQUENCE   240 AA;  24407 MW;  9FD9184ACC1221D8 CRC64;

  Query Match             6.4%;  Score 197.5;  DB 1;  Length 240;
  Best Local Similarity   31.6%;  Pred. No. 3.3e-07;
  Matches   78;  Conservative   32;  Mismatches  102;  Indels   35;  Gaps   12;

Qy         92 SSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQ-- 149
              |:| | |:|   ||  :    :   |  ||||:||:|    |    |: | :  |||
Db          7 STNAEALKL---APGLSGSAWVLAREQFAGRGRRGREWVMPAGN---FA-GTLVLRPQGG 59

Qy        150 -YELGSLSPVAAVACRRALSRLGL------KTQIKWPNDLVVGRDKLGGILIETVRTG-G 201
                 || |||:|   |   |||      :  ||||||:::   |: |||:|:  :| |
Db         60 ALAAAQLSFVAALALYDA---LGLACGPAARLAIKWPNDVLLNGGKVAGILLESSGSGPG 116

Qy        202 KTVAVVGIGINFVLPKEVENAASVQ-SLFQTASRRGNADAAV-------LLETLLAELDA 253
                ||||:|        :| :|:       | :|    ||        ||    |   |
Db        117 VQAVAVGIGVNLA---GAPDAGAVEPGATPPVSVQGETGHAVDPEEFLDLLAPAFARWQA 173

Qy        254 VLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQ 313
              |  |   ||||    : |     |: : : | |    |  :|:|  | | |    |:|
Db        174 QLDTY---GFAPIRNAWLARAARLGEPI-IARTGTAESHGIFEGIDDSGALILRGPAGRQ 229

Qy        314 TVVSGEI 320
                : : |:
Db        230 VIPAAEV 236


RESULT 7
YACB_BACSU
ID   YACB_BACSU     STANDARD;      PRT;   233 AA.
AC   P37564;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yacB.
GN   YACB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             4.3%;  Score 133;  DB 1;  Length 233;
  Best Local Similarity   31.7%;  Pred. No. 0.013;
  Matches   38;  Conservative   19;  Mismatches   49;  Indels   14;  Gaps    2;

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
                |:   | :|:| |:||  ||: :     |  :||  ||| |   : ::  |:|| | ||
Db         95 GLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGI 154

Qy        481 HLMKESLAVRTANLNR--------HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLK 532
                 :  |:|  | | | |          ||       | :|: ||::    | |  :  |:|
Db        155 TISTEALYSRAAKLPRIEITRPDNIIGK------NTVSAMQSGILFGYVGQVEGIVKRMK 208


RESULT 8
OMPB_RICCN
ID   OMPB_RICCN     STANDARD;      PRT;  1655 AA.
AC   Q9KKA3; Q9KK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
```

```
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       SIMILARITY).
CC   -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE008659; AAL03623.1; -.
DR   EMBL; AF123721; AAF34124.1; -.
DR   EMBL; AF123726; AAF34129.1; -.
DR   EMBL; AF149110; AAD39533.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall; Complete proteome.
FT   CHAIN         1   1334       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1335   1655       32 KDA BETA PEPTIDE.
FT   VARIANT      61     61       P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      75     75       G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      78     78       K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     251    251       V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     413    413       N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     959    959       I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     988    988       A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    1139   1139       R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT   CONFLICT    353    354       KD -> GH (IN REF. 3).
FT   CONFLICT    776    776       F -> S (IN REF. 3).
FT   CONFLICT   1159   1159       E -> D (IN REF. 3).
FT   CONFLICT   1177   1177       G -> S (IN REF. 3).
FT   CONFLICT   1492   1492       H -> R (IN REF. 3).
SQ   SEQUENCE   1655 AA;  168342 MW;  E49E19377D5FCE37 CRC64;

  Query Match             4.2%;  Score 130.5;  DB 1;  Length 1655;
  Best Local Similarity   22.0%;  Pred. No. 0.24;
  Matches  102;  Conservative   63;  Mismatches  187;  Indels  111;  Gaps   21;

Qy        181 DLVVGRDKLGGILIETVR---------TGGKTVAVVGIGINFVLPKEVENAASVQSLFQT 231
              |: |  |   | :   |           |||: : | ||  |     :|| :| :
Db        101 DVTVREDTTLGFITNVVNNANHFNLMLNAGKTLTITGQGITNVQAAATKNANNVVAQVNN 160

Qy        232 ASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                  :  | |       : :|         | |     :  ||    || |:| |   :

Db        161 GAAIDNNDLQG-----VGRIDC--------GAAASTLVFNLANPTTQKAPLILGDNAVIV 207

Qy        292 EG-------------------TVKGV---DGQGVLHLETAEGKQTVVSGEISLRSDDRP 328
                |                   ||  :   ||||::    |:  |     ::|:::
Db        208 NGANGTLNVTNGFIKVSSKSFATVNVINIGDGQGIMFNTDADNVNT-----LNLQANGAT 262

Qy        329 VSVPKRRDSERFLLLD-------------GGNSRLKWAWVENGTFATVGSAPYR---DL 371
              ::      : | :||              |||  :     :|  | |  |        :
Db        263 ITFNGTDGTGRLVLLSKNAAATDFNVTGSLGGNLK---GIIEFNTVAVNGQLKANAGANA 319

Qy        372 SPLGAEWAEKVDGNVRIVGCAV-CGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPE 430
              : :|       :|  |  |  |      | | :  |:  | : :  || |:|  | :||
Db        320 AVIGTN-----NGAGRAAGFVVSVDNGKVATIDGQVYAK-DMVIQSANAVGQVN-FRHIV 372

Qy        431 EHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDA---LTDDGHYLGGTIMPGFHLMKESL 487
              : |:|      |:  |  |  |     |:| ::    || |:     |:|    :  :
Db        373 DVGTD-----GTTAFKTAASKV-----AITQNSNFGTTDFGNLAAQIIVPNTMTLNGNF 421

Qy        488 AVRTANLNRHAGKRYPFPTTTGNAVASGMMD---AVCGSVMMMHGRLKEKTGAGKPVDVI 544
                :|    ||           :||   |   ||  ::  :     | :|||    |
Db        422 TGDASNPGNTAG---VITFDANGTLASASADANVAVTNNITAI-----EASGAGV---VQ 470

Qy        545 ITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG 587
              ::|  ||::          | :  ::||  ||:| :|  |  ||
Db        471 LSGTHAAELRLG-----NAGSVFKLADGTVINGKVNQTALVGG 508


RESULT 9
VHSJ_LAMBD
ID   VHSJ_LAMBD     STANDARD;      PRT;  1132 AA.
AC   P03749;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   01-AUG-1992 (Rel. 23, Last annotation update)
DE   Host specificity protein J.
GN   J.
OS   Bacteriophage lambda.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   Lambda-like viruses.
OX   NCBI_TaxID=10710;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=83189071; PubMed=6221115;
RA   Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT   "Nucleotide sequence of bacteriophage lambda DNA.";
RL   J. Mol. Biol. 162:729-773(1982).
CC   -!- FUNCTION: GENE J PROTEIN MAKES UP THE DISTAL FIBER OF THE PHAGE
CC       TAIL. IT IS RESPONSIBLE FOR ADSORPTION TO THE HOST DURING
CC       INFECTION AND DETERMINES HOST SPECIFICITY. IT SPECIFICALLY BINDS
CC       TO THE LAMBDA RECEPTOR PROTEIN OF E.COLI K12.
CC   -!- MISCELLANEOUS: UNDER THE ACTION OF THE GENE G, T, H, M, L, K, AND
CC       I PROTEINS, GENE J PROTEIN SERVES AS THE INITIATOR OF TAIL
CC       POLYMERIZATION. THERE ARE 2-4 COPIES OF PROTEIN J PER MATURE
CC       PHAGE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J02459; AAA96553.1; -.
DR   PIR; A04374; QSBPL.
DR   InterPro; IPR003961; FN_III.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 1.
KW   Phage recognition; Fiber protein.
SQ   SEQUENCE   1132 AA;  124422 MW;  53D4AC9044CF3C05 CRC64;

  Query Match             4.0%;  Score 122;  DB 1;  Length 1132;
  Best Local Similarity   20.9%;  Pred. No. 0.61;
```

```
  Matches  130;  Conservative   70;  Mismatches  164;  Indels  258;  Gaps   33;

Qy        154 SLSPVAAVAC--RRALSRLGL---KTQIKWPN--DLVVGRDKL---GGILIET------- 196
              ::: : |  |  |    | ||   ||::      |  || : |    | :||
Db        429 NVTKMDAFGCTSRGQAHRAGLWLIKTELLETQTVDFSVGAEGLRHVPGDVIEICDDDYAG 488

Qy        197 VRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLL 256
              : |||: :||        | :|:   :|  :|      ||   :| ::::    | | :
Db        489 ISTGGRVLAVNSQTRTLTLDREITLPSSGTALISLVDGSGN-PVSVEVQSV---TDGVKV 544

Qy        257 QYAR--DGFAPFVAEYQA-------------------ANRDHGKAVLLLR---------- 285
              : :|  ||    ||||                    | |   |:  ::
Db        545 KVSRVPDG----VAEYSVWELKLPTLRQRLFRCVSIRENDDGTYAITAVQHVPEKEAIVD 600

Qy        286 -----DGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGE--------------------- 319
                  |||    ||| ||    | || |||   |  |||
Db        601 NGAHFDGEQ--SGTVNGVTPPAVQHL-TAE--VTADSGEYQVLARWDTPKVVKGVSFLLR 655

Qy        320 ISLRSDD---RPVSVPKRRDSE-RFLLLDGGNSRLKW----AWVENGTFATVG------S 365
              ::: :||   | ||  :  ::  ||  |  || ||      || : |  |:|       :
Db        656 LTVTADDGSERLVSTARTTETTYRFTQLALGNYRLTVRAVNAWGQQGDPASVSFRIAAPA 715

Qy        366 APYR-DLSP-------------------LGAEWAEK-------VDGNVRIVGCAVC---- 394
              || | :|:|                   ::||       |: : | :| |:
Db        716 APSRIELTPGYFQITATPHLAVYDPTVQFEFWFSEKQIADIRQVETSTRYLGTALYWIAA 775

Qy        395 ------------------------------------------GEFKKAQVQEQLARKI 410
                                                       |:  :: : ::|  |:
Db        776 SINIKPGHDYYFYIRSVNTVGKSAFVEAVGRASDDAEGYLDFFKGKITESHLGKELLEKV 835

Qy        411 EWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGH 470
              |    :|  |      :   :  ||:| ||:                || ::   |  |
Db        836 ELTEDNASRL---EEFSKEWKDASDKW-NAM---------------WAVKIEQTKDGKH 875

Qy        471 YLGGTIMPGFHLMKES----------LAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAV 520
              |: |  :     |:::          | | | ::   |   |     ||
Db        876 YVAGIGLS----MEDTEEGKLSQFLVAANRIAFIDPANGNETPMFVAQGN---------- 921

Qy        521 CGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAF--------LAEN------- 565
                 : |    ||  |          || ||         ||||          |:|
Db        922 --QIFMNDVFLKRLTAP------TITSGGN-------PPAFSLTPDGKLTAKNADISGSV 966

Qy        566 --------TVRVADNLVIHGLL 579
                     | :|:|  |:| |
Db        967 NANSGTLSNVTIAENCTINGTL 988


RESULT 10
BPL1_HUMAN
ID   BPL1_HUMAN     STANDARD;      PRT;   726 AA.
AC   P50747; Q99451;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Biotin--protein ligase (EC 6.3.4.-) (Biotin apo-protein ligase)
DE   [Includes: Biotin--[methylmalonyl-CoA-carboxyltransferase] ligase
DE   (EC 6.3.4.9); Biotin--[propionyl-CoA-carboxylase [ATP-hydrolyzing]]
DE   ligase (EC 6.3.4.10) (Holocarboxylase synthetase) (HCS);
DE   Biotin--[methylcrotonoyl-CoA-carboxylase] ligase (EC 6.3.4.11);
DE   Biotin--[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)].
GN   HLCS.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND VARIANT HCS PRO-237.
RC   TISSUE=Liver;
RX   MEDLINE=95144167; PubMed=7842009;
RA   Suzuki Y., Aoki Y., Ishida Y., Chiba Y., Iwamatsu A., Kishino T.,
RA   Niikawa N., Matsubara Y., Narisawa K.;
RT   "Isolation and characterization of mutations in the human
RT   holocarboxylase synthetase cDNA.";
RL   Nat. Genet. 8:122-128(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Ohira M., Seki N., Nagase T., Suzuki E., Nomura N., Ohara O.,
RA   Hattori M., Sakaki Y., Eki T., Murakami Y., Saito T., Ichikawa H.,
RA   Ohki M.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Shibuya K., Kudoh J., Minoshima S., Kawasaki K., Nakatoh E.,
RA   Shintani A., Asakawa S., Shimizu N.;
RT   "Genomic sequencing of 1.2-Mb region on human chromosome 21q22.2.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A., AND VARIANTS HCS P-237; W-508; M-550 AND S-581.
RX   PubMed=11735028;
RA   Yang X., Aoki Y., Li X., Sakamoto O., Hiratsuka M., Kure S.,
RA   Taheri S., Christensen E., Inui K., Kubota M., Ohira M., Ohki M.,
RA   Kudoh J., Kawasaki K., Shibuya K., Shintani A., Asakawa S.,
RA   Minoshima S., Shimizu N., Narisawa K., Matsubara Y., Suzuki Y.;
RT   "Structure of human holocarboxylase synthetase gene and mutation
RT   spectrum of holocarboxylase synthetase deficiency.";
RL   Hum. Genet. 109:526-534(2001).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
RN   [6]
RP   SEQUENCE OF 1-92 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98163741; PubMed=9503011;
RA   Dahmane N., Ait Ghezala G., Gosset P., Chamoun Z.,
RA   Dufresne-Zacharia M.C., Lopes C., Rabatel N., Gassanova-Maugenre S.,
RA   Chettouh Z., Abramowski V., Fayet E., Yaspo M.L., Korn B.,
RA   Blouin J.L., Lehrach H., Poustka A., Antonarakis S.E., Sinet P.M.,
RA   Creau N., Delabar J.M.;
RT   "Transcriptional map of the 2.5-Mb CBR-ERG region of chromosome 21
RT   involved in down syndrome.";
RL   Genomics 48:12-23(1998).
RN   [7]
RP   VARIANT HCS PRO-237.
RX   MEDLINE=96125714; PubMed=8541348;
RA   Aoki Y., Suzuki Y., Sakamoto O., Li X., Takahashi K., Ohtake A.,
RA   Sakuta R., Ohura T., Miyabayashi S., Narisawa K.;
RT   "Molecular analysis of holocarboxylase synthetase deficiency: a
RT   missense mutation and a single base deletion are predominant in
RT   Japanese patients.";
RL   Biochim. Biophys. Acta 1272:168-174(1995).
RN   [8]
RP   VARIANT HCS ASN-571.
RX   MEDLINE=96414311; PubMed=8817339;
RA   Dupuis L., Leon-Del-Rio A., Leclerc D., Campeau E., Sweetman L.,
RA   Saudubray J.M., Herman G., Gibson K.M., Gravel R.A.;
RT   "Clustering of mutations in the biotin-binding region of
RT   holocarboxylase synthetase in biotin-responsive multiple carboxylase
RT   deficiency.";
RL   Hum. Mol. Genet. 5:1011-1016(1996).
RN   [9]
RP   VARIANTS HCS PRO-237 AND MET-550.
RX   MEDLINE=98057254; PubMed=9396568;
```

```
RA   Aoki Y., Suzuki Y., Li X., Sakamoto O., Chikaoka H., Takita S.,
RA   Narisawa K.;
RT   "Characterization of mutant holocarboxylase synthetase (HCS): a Km for
RT   biotin was not elevated in a patient with HCS deficiency.";
RL   Pediatr. Res. 42:849-854(1997).
RN   [10]
RP   VARIANT HCS GLU-333; ILE-462; ASN-571; SER-581 AND THR-610 DEL.
RX   MEDLINE=99204600; PubMed=10190325;
RA   Aoki Y., Li X., Sakamoto O., Hiratsuka M., Akaishi H., Xu L.,
RA   Briones P., Suormala T., Baumgartner E.R., Suzuki Y., Narisawa K.;
RT   "Identification and characterization of mutations in patients with
RT   holocarboxylase synthetase deficiency.";
RL   Hum. Genet. 104:143-148(1999).
CC   -!- FUNCTION: POSTTRANSLATIONAL MODIFICATION OF SPECIFIC PROTEIN BY
CC       ATTACHMENT OF BIOTIN. ACTS ON VARIOUS CARBOXYLASES SUCH AS ACETYL-
CC       COA-CARBOXYLASE, PYRUVATE CARBOXYLASE, PROPIONYL COA CARBOXYLASE,
CC       AND 3-METHYLCROTONYL COA CARBOXYLASE.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[methylmalonyl-CoA:pyruvate
CC       carboxyltransferase] = AMP + diphosphate + [methylmalonyl-
CC       CoA:pyruvate carboxyltransferase].
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[propanoyl-CoA:carbon-
CC       dioxide ligase (ADP-forming)] = AMP + diphosphate + [propanoyl-
CC       CoA:carbon-dioxide ligase (ADP-forming)].
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[3-methylcrotonoyl-
CC       CoA:carbon-dioxide ligase (ADP-forming)] = AMP + diphosphate + [3-
CC       methylcrotonoyl-CoA:carbon-dioxide ligase (ADP-forming)].
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and mitochondrial.
CC   -!- TISSUE SPECIFICITY: MOSTLY EXPRESSED IN MUSCLE, PLACENTA, IN
CC       LESSER EXTENT IN THE BRAIN, KIDNEY, PANCREAS, LIVER AND LUNG.
CC   -!- DISEASE: DEFECTS IN HLCS ARE THE CAUSE OF HOLOCARBOXYLASE
CC       SYNTHETASE DEFICIENCY (HCS) (OR BIOTIN-RESPONSIVE MULTIPLE
CC       CARBOXYLASE DEFICIENCY), AN AUTOSOMAL RECESSIVE DISORDER
CC       CHARACTERIZED BY METABOLIC KETOACIDOSIS, HYPERAMMONEMIA, EXCRETION
CC       OF ABNORMAL ORGANIC ACID METABOLITES, AND DERMATITIS. CLINICAL AND
CC       BIOLOCHEMICAL SYMPTOMS IMPROVE DRAMATICALLY WITH ADMINISTRATION OF
CC       BIOTIN.
CC   -!- SIMILARITY: WITH E.COLI BIRA AND OTHER EUKARYOTIC BIOTIN--PROTEIN
CC       LIGASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D23672; BAA04902.1; -.
DR   EMBL; D87328; BAA13332.1; -.
DR   EMBL; AP000697; BAA89434.1; -.
DR   EMBL; AP000703; BAA89434.1; JOINED.
DR   EMBL; AP000701; BAA89434.1; JOINED.
DR   EMBL; AP000698; BAA89434.1; JOINED.
DR   EMBL; AB063285; BAB68550.1; -.
DR   EMBL; AP001727; BAA95511.1; -.
DR   EMBL; AP001726; BAA95510.1; -.
DR   EMBL; AJ001864; CAA05056.1; -.
DR   Genew; HGNC:4976; HLCS.
DR   MIM; 253270; -.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase; Multifunctional enzyme; Disease mutation.
FT   SIMILAR     563    609       SOME, TO AVIDIN.
FT   VARIANT     237    237       L -> P (IN HCS).
FT                                /FTId=VAR_005084.
FT   VARIANT     333    333       V -> E (IN HCS; <10% ACTIVITY).
FT                                /FTId=VAR_009196.
FT   VARIANT     462    462       T -> I (IN HCS; <10% ACTIVITY).
FT                                /FTId=VAR_009197.
FT   VARIANT     508    508       R -> W (IN HCS).
FT                                /FTId=VAR_013009.
FT   VARIANT     550    550       V -> M (IN HCS).
FT                                /FTId=VAR_009198.
FT   VARIANT     571    571       D -> N (IN HCS; ALMOST NO ACTIVITY).
FT                                /FTId=VAR_009199.
FT   VARIANT     581    581       G -> S (IN HCS; <10% ACTIVITY).
FT                                /FTId=VAR_009200.
FT   VARIANT     610    610       MISSING (IN HCS; 14% OF ACTIVITY).
FT                                /FTId=VAR_009201.
FT   CONFLICT    558    558       E -> K (IN REF. 2).
SQ   SEQUENCE   726 AA;  80759 MW;  855B8E52106D675F CRC64;

  Query Match             3.9%;  Score 121.5;  DB 1;  Length 726;
  Best Local Similarity   23.0%;  Pred. No. 0.37;
  Matches   58;  Conservative   45;  Mismatches  100;  Indels   49;  Gaps   13;

Qy        112 ICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRP-QYELGSLSP-------VAAVAC 163
              |  :   |::|:||  |   |    :| | : :    :    | : :||   |       || |
Db        496 IVIAARQTEGKGRGGNVWLSPVG-CALSTL--LISIPLRSQLGQRIPFVQHLMSVAVVEA 552

Qy        164 RRALSRL-GLKTQIKWPNDLVVG-RDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVEN 221
               |::    :  ::|||||:      |:||:|: :   |     ::| | |       | |
Db        553 VRSIPEYQDINLRVKWPNDIYYSDLMKIGGVLVNSTLMGETFYILIGCGFN------VTN 606

Qy        222 A---ASVQSLFQTASRRGNADAAVL-LETLLAELDAVLLQYARDGFAPFVAEYQAANRDH 277
              :     :  |    :::  |:   |  : |:| :  ||          : |:|    |
Db        607 SNPTICINDLITEYNKQHKAELKPLRADYLIARVVTVL--------EKLIKEFQ----DK 654

Qy        278 GKAVLL-------LRDGETVFEG-------TVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              |   :|       :  |: |  |       :: |:|  | | :    |:   |  : :
Db        655 GPNSVLPLYYRYWVHSGQQVHLGSAEGPKVSIVGLDDSGFLQVHQEGGEVVTVHPDGNSF 714

Qy        324 SDDRPVSVPKRR 335
                | : :||||
Db        715 DMLRNLILPKRR 726


RESULT 11
GLTD_ECOLI
ID   GLTD_ECOLI     STANDARD;      PRT;   471 AA.
AC   P09832;
DT   01-MAR-1989 (Rel. 10, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Glutamate synthase [NADPH] small chain (EC 1.4.1.13) (Glutamate
DE   synthase beta subunit) (NADPH-GOGAT) (GLTS beta chain).
GN   GLTD OR ASPB OR B3213.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   STRAIN=K12;
RX   MEDLINE=88152492; PubMed=3326786;
RA   Oliver G., Gosset G., Sanchez-Pescador R., Lozoya E., Ku L.M.,
RA   Flores N., Becerril B., Valle F., Bolivar F.;
RT   "Determination of the nucleotide sequence for the glutamate synthase
RT   structural genes of Escherichia coli K-12.";
RL   Gene 60:1-11(1987).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
```

```
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [3]
RP   SEQUENCE OF 1-12.
RC   STRAIN=K12 / EMG2;
RX   MEDLINE=97443975; PubMed=9298646;
RA   Link A.J., Robison K., Church G.M.;
RT   "Comparing the predicted and observed properties of proteins encoded
RT   in the genome of Escherichia coli K-12.";
RL   Electrophoresis 18:1259-1313(1997).
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=89098858; PubMed=2643092;
RA   Gosset G., Merino E., Recillas F., Oliver G., Becerril B., Bolivar F.;
RT   "Amino acid sequence analysis of the glutamate synthase enzyme from
RT   Escherichia coli K-12.";
RL   Protein Seq. Data Anal. 2:9-16(1989).
CC   -!- CATALYTIC ACTIVITY: 2 L-glutamate + NADP(+) = L-glutamine + 2-
CC       oxoglutarate + NADPH.
CC   -!- COFACTOR: BINDS A 4FE-4S CLUSTER.
CC   -!- PATHWAY: NITROGEN METABOLISM, GLUTAMATE BIOSYNTHESIS.
CC       THE CATALYZED REACTION BRINGS TOGETHER THE NITROGEN AND
CC       CARBON METABOLISM.
CC   -!- SUBUNIT: AGGREGATE OF 4 CATALYTICAL ACTIVE HETERODIMERS,
CC       CONSISTING OF A LARGE AND A SMALL SUBUNIT.
CC   -!- MISCELLANEOUS: GLUTAMINE BINDS TO THE LARGE SUBUNIT AND TRANSFERS
CC       THE AMIDO GROUP TO 2-OXO-GLUTAMATE THAT APPARENTLY BINDS TO THE
CC       SMALL SUBUNIT.
CC   -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC       'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC   -!- SIMILARITY: STRONG, TO E.COLI AEGA AND YGFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M18747; AAA23905.1; -.
DR   EMBL; U18997; AAA58015.1; -.
DR   EMBL; AE000400; AAC76245.1; -.
DR   PIR; B29617; B29617.
DR   SWISS-2DPAGE; P09832; COLI.
DR   ECO2DBASE; F050.4; 6TH EDITION.
DR   EcoGene; EG10404; gltD.
DR   InterPro; IPR000759; Adrndx_reductase.
DR   InterPro; IPR001327; FAD_pyr_redox.
DR   InterPro; IPR000205; NAD_binding.
DR   Pfam; PF00070; pyr_redox; 1.
DR   PRINTS; PR00419; ADXRDTASE.
DR   ProDom; PD000139; FAD_pyr_redox; 1.
DR   TIGRFAMs; TIGR01318; gltD_gamma_fam; 1.
KW   Oxidoreductase; Glutamate biosynthesis; Iron-sulfur; 4Fe-4S; NADP;
KW   Complete proteome.
FT   INIT_MET      0      0
FT   METAL         93     93       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL         97     97       IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL         103    103      IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   METAL         107    107      IRON-SULFUR (4FE-4S) (POTENTIAL).
FT   CONFLICT      37     50       GQAKAQADRCLSCG -> ARPKRRLTAACRAA (IN
FT                                 REF. 1).
FT   CONFLICT      122    122      E -> K (IN REF. 1).
FT   CONFLICT      173    173      V -> C (IN REF. 1).
FT   CONFLICT      256    269      VYAALPFLIANTKQ -> CTQRCRSSSPTPNS (IN
FT                                 REF. 1).
FT   CONFLICT      311    312      KH -> ND (IN REF. 1).
FT   CONFLICT      375    399      GRRRAEIVAGSEHIVPADAVIMAFG -> ASPRGDRCRFRT
FT                                 YRTGRCGDHGVW (IN REF. 1).
SQ   SEQUENCE   471 AA;  51884 MW;  C4070F7509C561A8 CRC64;

  Query Match             3.9%;  Score 121;  DB 1;  Length 471;
  Best Local Similarity   22.8%;  Pred. No. 0.23;
  Matches  102;  Conservative   44;  Mismatches  125;  Indels  176;  Gaps   24;

Qy        139 FSFGWVFDRPQY----ELGSLSPV-----AAVACRRALSRLGLKTQIKWPNDLVVGR-DK 188
              |  ||   ||       : |    :     | :||   |:| |:|        :|  | :
Db        130 FEMGW---RPDMSGVKQTGKKVAIIGAGPAGLACADVLTRNGVKA-------VVFDRHPE 179

Qy        189 LGGILI---------ETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNAD 239
              :||:|          : | |  : :   |:|| | |  ||
Db        180 IGGLLTFGIPAFKLEKEVMTRRREI-FTGMGIEFKLNTEV------------------G 219

Qy        240 AAVLLETLLAELDAVLL-----QYARDGF-----------APFVAEYQAANRDHGKAVLL 283
                | |: ||:: ||| |     |  | |             ||:     ||        |
Db        220 RDVQLDDLLSDYDAVFLGVGTYQSMRGGLENEDADGVYAALPFL----IANTKQ-----L 270

Qy        284 LRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSG---------EISLRSDDRPVSVPKR 334
              :  |||  |  |            : |||: || |           |:|   : |:   |
Db        271 MGFGETRDEPFV-----------SMEGKRVVVLGGGDTAMDCVRTSVRQGAKHVTCAYR 318

Qy        335 RDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAV- 393
              || |    : |    :| | |  |         :: ||| |    |:|| :: | :
Db        319 RDEEN---MPGSRREVKNAREEGVEFKF-------NVQPLGIE----VNGNGKVSGVKMV 364

Qy        394 ---CGE--FKKAQVQEQLARKIEWLPSSA--QALGIRNH-----YRHPEE---------- 431
                 ||   |  :  | :|     :|: |   | | | |      :|  |
Db        365 RTEMGEPDAKGRRRAEIVAGSEHIVPADAVIMAFGFRPHNMEWLAKHSVELDSQGRIIAP 424

Qy        432 HGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT 491
               |||  |    : |:                          || |: |  |:  ::|
Db        425 EGSDNAFQTSNPKIFA-----------------------GGDIVRGSDLVVTAIA--- 456

Qy        492 ANLNRHAGKRYPFPTTTGNAVASGMMD 518
                          |   | |:|:
Db        457 ----------------EGRKAADGIMN 467


RESULT 12
OMPB_RICJA
ID   OMPB_RICJA     STANDARD;      PRT;  1656 AA.
AC   O06653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia japonica.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=35790;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YH;
RA   Uchiyama T.;
RT   "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT   japonica.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       SIMILARITY).
CC   -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB003681; BAA20138.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall.
FT   CHAIN         1   1338       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1339   1656       32 KDA BETA PEPTIDE.
FT   DOMAIN      528    533       POLY-GLY.
SQ   SEQUENCE   1656 AA;  168097 MW;  3132A69C9DD5999F CRC64;

  Query Match             3.8%;  Score 118;  DB 1;  Length 1656;
  Best Local Similarity   22.6%;  Pred. No. 1.9;
  Matches  100;  Conservative   52;  Mismatches  171;  Indels  120;  Gaps   21;

Qy        201 GKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETL--LAELDAVLLQY 258
              |||: : | ||  |      || :: : |       |  ||:    |  |  :|
Db        130 GKTLTITGQGITNVQSAATHNAQNIVAKF-------NGGAAIANNDLSGLGTIDF----- 177

Qy        259 ARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEG--------------------TVKGV 298
              | |     :  ||    || |:| |   :  |                       ||| :
Db        178 ---GAAASTLVFDLANPTTQKAPLILADNALIVNGANGTLNVTNGFIQVSDKSFATVKAI 234

Qy        299 ---DGQGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLD---------- 344
                 |||| :    |       :  ::|::    ::       : | :||
Db        235 NIGDGQGFMFNTNATN-----ANALNLQAGGTTINFNGTDGTGRLVLLSKNGAATDFNVT 289

Qy        345 ---GGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQ 401
                 ||| :     :|  | | |     :  | |      :|  |  |  |  :  ||
Db        290 GSLGGNLK---GIIELNTVAINGQL-IANAGPANAVIGTN-NGAGRAAGFVVSVDNGKAA 344

Qy        402 VQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVV-----SC 456
                :     : :  || | |  | :||  : | |      |:  |   | :|      :
Db        345 TIDGQVYAKDMVIQSANANGQVN-FRHIVDVGID------GTTAFKTAASIVAITQNSNF 397

Qy        457 GTA--------VTV-DALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTT 507
              ||         ||| | :|  |:: |    ||           || :   |
Db        398 GTTDFGNLAAQVTVPDTMTLTGNFTGDANNPG----------NTAGVITFA--------- 438

Qy        508 TGNAVASGMMD---AVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAE 564
                 :||   |   ||  ::  :     | :| |    | ::|   |::          |
Db        439 ANGTLASASADANVAVTNNITAI-----EASGVGV---VQLSGTHTAELRLG-----NAG 485

Qy        565 NTVRVADNLVIHGLLNLIAAEGG 587
              :  ::||  ||:| :|     ||
Db        486 SVFKLADGTVINGKVNQTVLVGG 508


RESULT 13
OMPB_RICRI
ID   OMPB_RICRI     STANDARD;      PRT;  1654 AA.
AC   Q53047;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia rickettsii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=783;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R;
RX   MEDLINE=92167802; PubMed=1724278;
RA   Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;
RT   "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT   rickettsii is encoded by an unusually long open reading frame:
RT   evidence for protein processing from a large precursor.";
RL   Mol. Microbiol. 5:2361-2370(1991).
RN   [2]
RP   SEQUENCE OF 279-1654 FROM N.A.
RC   STRAIN=R;
RX   MEDLINE=90136087; PubMed=2515418;
RA   Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT   "Cloning, expression and sequence analysis of the gene encoding the
RT   120 kD surface-exposed protein of Rickettsia rickettsii.";
RL   Mol. Microbiol. 3:1579-1586(1989).
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X16353; CAA34403.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall.
FT   CHAIN         1   1333       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1334   1654       32 KDA BETA PEPTIDE.
FT   DOMAIN     1181   1188       POLY-THR.
SQ   SEQUENCE   1654 AA;  168184 MW;  D7AB70FB7087F618 CRC64;

  Query Match             3.7%;  Score 112.5;  DB 1;  Length 1654;
  Best Local Similarity   22.2%;  Pred. No. 4.7;
  Matches  106;  Conservative   54;  Mismatches  175;  Indels  143;  Gaps   23;

Qy        182 LVVGRDKLGGILIETVR---------TGGKTVAVVGIGINFVLPKEVENAASVQSLFQTA 232
              : |  |   | :  |             |||: : | |:        :|| :|   |
Db        102 VTVSEDTTLGFITNVVHNAHSFNLTLNAGKTLTITGQGVTNAQAAATKNAQNVVVQFNNG 161

Qy        233 SRRGNADAAVLLETLLAELDAVLLQYARDGF-AP-FVAEYQAANRDHGKAVLLLRDGETV 290
              :  | |               |    |  | ||     :  ||    || |:| |   :
Db        162 AAIDNND--------------LKGVGRIDFGAPASTLVFNLANPTTQKAPLILGDNAVI 206

Qy        291 FEG--------------------TVKGV---DGQGVLHLETAEGKQTVVSGEISLRSDDR 327
                |                    |||  :   ||||::    |    |      ::|::
Db        207 ANGVNGTLNVTNGFIQVSNKSFATVKAINIADGQGIIFNTDANNANT-----LNLQAGGT 261

Qy        328 PVSVPKRRDSERFLLLD--------------GGNSRLKWAWVENGTFATVGSAPYRDLSP 373
              ::       : | :||               |||  :     :|  | |  |
Db        262 TINFTGTDGTGRLVLLSKHAAATNFNITGSLGGNLK---GVIEFNTVAVDGQL------T 312

Qy        374 LGAEWAEKV----DGNVRIVGCAV-CGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRH 428
              |  |  |     :|  |  |  |      | | :  |:  | : :  || | |  | :||
Db        313 ANAGAANAVIGTNNGAGRAAGFVVSVDNGKVATIDGQVYAK-DMVIQSANATGQVN-FRH 370

Qy        429 PEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVD------------------ALTDDG 469
              : |:|       |:  |   |  |  |       :| |                    |:|  |
Db        371 IVDVGAD------GTTAFKTAASKV-----TITQDSNFGNTDFGNLAAQIKVPNAITLTG 419

Qy        470 HYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHG 529
              :: |    ||      :  | | : |           |  :| |   : ||  ::  :
Db        420 NFTGDASNPG-----NTAGVITFDAN----------GTLESASADANV-AVTNNITAI-- 461

Qy        530 RLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG 587
                | :|||    | ::|  ||::          | :  ::||  ||:| :|  |  ||
Db        462 ---EASGAGV---VQLSGTHAAELRLG-----NAGSIFKLADGTVINGKVNQTALVGG 508


RESULT 14
HYPF_AZOCH
```

```
ID   HYPF_AZOCH     STANDARD;      PRT;   776 AA.
AC   Q43950;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Hydrogenase maturation protein hypF.
GN   HYPF OR HUPY.
OS   Azotobacter chroococcum mcd 1.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Azotobacter.
OX   NCBI_TaxID=355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93252280; PubMed=8486288;
RA   Tibelius K.H., Du L., Tito D., Stejskal F.;
RT   "The Azotobacter chroococcum hydrogenase gene cluster: sequences and
RT   genetic analysis of four accessory genes, hupA, hupB, hupY and
RT   hupC.";
RL   Gene 127:53-61(1993).
RN   [2]
     ERRATUM.
A    Tibelius K.H., Du L., Tito D., Stejskal F.;
RL   Gene 134:141-141(1993).
CC   -!- FUNCTION: INVOLVED IN THE HYDROGENASE MATURATION PROCESS (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE HYPF FAMILY.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE
CC       ACYLPHOSPHATASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L00674; AAA22134.1; ALT_INIT.
DR   HSSP; P41500; 2ACY.
DR   InterPro; IPR001792; Acylphosphatase.
DR   InterPro; IPR004421; HypF.
DR   InterPro; IPR000666; Sua5_yciO_yrdC.
DR   Pfam; PF00708; Acylphosphatase; 1.
DR   Pfam; PF01300; Sua5_yciO_yrdC; 1.
DR   ProDom; PD001884; Acylphosphatase; 1.
DR   ProDom; PD002209; Sua5_yciO_yrdC; 1.
DR   TIGRFAMs; TIGR00143; hypF; 1.
DR   PROSITE; PS00150; ACYLPHOSPHATASE_1; 1.
KW   Zinc-finger.
FT   DOMAIN       1     93       ACYLPHOSPHATASE-LIKE.
FT   ZN_FING     113    138      C4-TYPE (POTENTIAL).
FT   ZN_FING     163    188      C4-TYPE (POTENTIAL).
SQ   SEQUENCE   776 AA;  82568 MW;  6ACDA50B8F90911B CRC64;

  Query Match             3.6%;  Score 110.5;  DB 1;  Length 776;
  Best Local Similarity   22.7%;  Pred. No. 2.5;
  Matches  156;  Conservative   61;  Mismatches  240;  Indels  231;  Gaps   39;

Qy         34 PQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFDA---EGLRELGERSGFQTALKHEC 90
              |  |  | :::|| :               |||  ||   | |      :||
Db         54 PAALAEFERRLPAEL-------------PPLARLDALQRENLPAGAHFAGF------AI 93

Qy         91 ASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECL----MFSF--GWV 144
                | ||        || |    |:  |         |:| |    |      :|
Db         94 VPSRDEAGGALAPPPDAAVCPACLEELFDP----HDRRWRHPFINCTHCGPRYSLIRRLP 149

Qy        145 FDRPQYELGSLS-------------------PVAAVACRRALSRLGLKTQIKWPNDLVV 184
              :|| |  |   :                  |:|  ||   |          |  |
Db        150 YDRAQTSLAGFALCPDCRHEYEDPADRRFHAQPIACPACGPRL----------WCED--A 197

Qy        185 GRDKLGG----ILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR----- 235
                ::| |    : : |:|  |: ||: |:|  | |  :  || :| :| :   ||
Db        198 AGNRLPGEPVELALATLRR-GEIVALRGVG-GFHLACDARNAGAV-ALLRRRKRRPAKPF 254

Qy        236 -------GNADAAV-LLETLLAELDA-----VLLQY--ARDGF-----APFVA------- 268
                     : :| |  | || | || |     |||:   | ||      || :|
Db        255 ALMAANPASLEALVELNETGLNELSAPAAPVVLLRRRPAADGLLAAEVAPDLAWLGVMLP 314

Qy        269 --------EYQAANRDHGKA--------VLLLR----DGETVFEGTVK------GVDGQG 302
                       ::|| |  | |        ||::      || :  | ::      |:
Db        315 HSPLHWLLFHEAAGRPAGTAWTAAPQDLVLVMTSANLSGEPLITGNLEAREKLAGITDLW 374

Qy        303 VLH-----------LETAEGKQTVVSGEISLRSDD--RPVSVPKRRDSERFLLLDGGNSR 349
              :||            |  | |:  ||     :||     |  :   |     | | |
Db        375 LLHDREILNRVDDSLVNALGRAPVV-----IRSGRGLAPQEIRLARSGPSVLALGG---- 425

Qy        350 LKWAWVENGTFATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARK 409
                ::|    | |   :  |||                        |:   |     | |
Db        426 ----QLKNAICLTQGDRAW--LSP------------------HIGDLHSADACRSLERT 460

Qy        410 IEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRN-------------------A 450
              :       : ||:|     ||    |   :   | ||:|                    |
Db        461 V---AQFGELLGVR-----PERIACDLHPDFFAS-RFAREHAERHGLPLYPVQHHHAHIA 511

Qy        451 CVVVSCGTAVTVDALTDDGHYLG--GTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTT 508
               |:   | |  |  |  ||  ||  | :  |     | | |          |:  | |
Db        512 AVLAEHGLAEPVLGLALDGFGLGADGRLRGG-----ELLKVAVDGC-AWLGELNPLPLPG 565

Qy        509 GNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL------PPAFL 562
              |:  |        |::  : ||      |:          |:| :|  |      ||:
Db        566 GDQAAREPWRMAAGALHAL-GR------GGEIATRFAAEPGSAAIARMLERGFNCPPSSS 618

Qy        563 AENTVRVADNLVIHGLLNLIAAEGGESE 590
              |      |      ||| |   :  |:|
Db        619 AGRLFDAA-----AGLLGLGERQRYEAE 641


RESULT 15
FRVX_ECOLI
ID   FRVX_ECOLI     STANDARD;      PRT;   356 AA.
AC   P32153;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative frv operon protein frvX.
GN   FRVX OR B3898.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=93347969; PubMed=8346018;
RA   Plunkett G. III, Burland V.D., Daniels D.L., Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. III. DNA sequence of the
RT   region from 87.2 to 89.2 minutes.";
RL   Nucleic Acids Res. 21:3391-3398(1993).
RN   [2]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=94290319; PubMed=8019415;
RA   Reizer J., Michotey V., Reizer A., Saier M.H. Jr.;
RT   "Novel phosphotransferase system genes revealed by bacterial genome
RT   analysis: unique, putative fructose- and glucoside-specific
RT   systems.";
RL   Protein Sci. 3:440-450(1994).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M42.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; L19201; AAB03031.1; -.
DR   EMBL; AE000465; AAC76880.1; -.
DR   PIR; S40842; S40842.
DR   MEROPS; M42.UPW; -.
DR   EcoGene; EG11862; frvX.
KW   Hydrolase; Complete proteome.
SQ   SEQUENCE   356 AA;  38733 MW;  8F399EFDA3ADE353 CRC64;

  Query Match             3.6%;  Score 110;  DB 1;  Length 356;
  Best Local Similarity   23.6%;  Pred. No. 1;
  Matches   95;  Conservative   54;  Mismatches  157;  Indels   96;  Gaps   23;

Qy        221 NAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKA 280
              |   :| | : ::  |  |   : : |:  |:  : :  ||  |||  :  |:    |
Db          2 NIELLQQLCEASAVSG--DEQEVRDILINTLEPCVNEITFDGLGSFVA--RKGNKGPKVA 57

Qy        281 VLLLRDGETVFEGTVKGVDGQGVLHLETAEG--KQTVVSGEISLRSDDRPVSVP------ 332
              |:   | |  |   |  :| | |   |  |   |::::  :::|:  : | :|
Db         58 VVGHMD-EVGF--MVTHIDESGFLRFTTIGGWWNQSMLNHRVTIRT-HKGVKIPGVIGSV 113

 y        333 -----KRRDSERFLLLD------GGNSR--LKWAWVENGTFATVGSAPYRDLSPLG--AE 377
                    :  :: |  |       | |||  ::   || | |          :||    |
Db        114 APHALTEKQKQQPLSFDEMFIDIGANSREEVEKRGVEIGNF----------ISPEANFAC 163

Qy        378 WAE-KVDGNV--RIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGS 434
              | | || |      :|||:  |  :     ::   :  : |  : :|:|      |
Db        164 WGEDKVVGKALDNRIGCAMMAELLQTVNNPEIT--LYGVGSVEEEVGLRGAQTSAEHIKP 221

Qy        435 DRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANL 494
              |                ||:   |||  |             :||   :|  |      |
Db        222 D---------------VVIVLDTAVAGD------------VPGIDNIKYPL-----KL 247

Qy        495 NRHAG-----KRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGG 549
               :  |     ||| ||      ||:    |   : :    :  ||||       : |||
Db        248 GQGPGLMLFDKRY-FPNQ--KLVAALKSCAAHNDLPLQFSTM--KTGATDGGRYNVMGGG 302

Qy        550 AAKVAEALPPAFLAENT--VRVADN----LVIHGLLNLIAAE 585
                ||  ||  :|  |:  :  ||      :| | |  : ||
Db        303 RPVVALCLPTRYLHANSGMISKADYEALLTLIRGFLTTLTAE 344


Search completed: June 24, 2003, 21:51:31
Job time : 12.2965 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 21.5927 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-22
Perfect score: 3080
Sequence: 1 MTVLKPSHWRVLAELADGLP.....LVIHGLLNLIAAEGGESEHT 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3080	100.0	592	2	H82031	probable biotin-[a
2	3024	98.2	592	2	B81009	BirA protein/Bvg a
3	305.5	9.9	321	2	D91241	hypothetical prote
4	305.5	9.9	321	2	A86089	hypothetical prote
5	304.5	9.9	321	1	BVECBF	biotin-[acetyl-CoA
6	287.5	9.3	320	2	E82337	birA bifunctional
7	281.5	9.1	312	2	A83112	BirA bifunctional
8	276.5	9.0	319	2	AG0457	biotin-[acetyl-CoA
9	269.5	8.8	320	2	AG0934	bifunctional prote
10	256.5	8.3	267	2	I40327	baf protein - Bord
11	244	7.9	474	2	D75285	birA bifunctional
12	232.5	7.5	317	2	B82637	bifunctional bioti
13	231.5	7.5	248	2	H83111	hypothetical prote
14	226.5	7.4	325	2	A69595	transcription repr
15	218.5	7.1	274	2	E87000	biotin apo-protein
16	218.5	7.1	325	2	AH1684	transcription regu
17	217.5	7.1	325	2	AH1312	transcription regu
18	215	7.0	323	2	E83860	biotin operon tran
19	214	6.9	266	2	H70979	probable birA prot
20	208.5	6.8	274	2	S74680	biotin-[acetyl-CoA
21	203	6.6	233	2	A70351	biotin [acetyl-CoA
22	198.5	6.4	242	2	A82637	conserved hypothet
23	197.5	6.4	240	2	B47751	biotin [acetyl-CoA
24	193.5	6.3	261	2	H69122	biotin acetyl-CoA
25	189.5	6.2	237	2	H75202	biotin operon repr
26	189	6.1	235	2	A71236	probable biotin-[a
27	183	5.9	273	2	E97293	probable transcrip
28	182.5	5.9	323	2	H89923	hypothetical prote
29	181.5	5.9	250	2	C87489	biotin protein lig
30	180.5	5.9	265	2	T36391	hypothetical prote
31	179	5.8	311	2	B95222	BirA bifunctional
32	179	5.8	311	2	A99086	bioti-[acetyl-CoA-
33	177.5	5.8	231	2	C72401	biotin-(acetyl-CoA
34	175.5	5.7	256	2	AD2734	birA bifunctional
35	175.5	5.7	256	2	F97515	birA protein (AF20
36	175	5.7	261	2	B87489	transcription acti
37	171.5	5.6	268	2	D71657	biotin-protein lig
38	169.5	5.5	282	2	B84224	biotin acetyl-CoA
39	169	5.5	323	2	D86847	biotin-[acetyl-CoA
40	166	5.4	250	2	G86862	biotin-[acetyl-CoA
41	166	5.4	274	2	AH2462	biotin acetyl-CoA
42	161	5.2	355	2	G84651	biotin holocarboxy
43	156	5.1	270	2	E97796	biotin-[acetyl-CoA
44	154.5	5.0	274	2	H86937	conserved hypothet
45	154.5	5.0	310	2	G86490	probable biotin ho

ALIGNMENTS

RESULT 1
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningiti
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C;Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.;
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z24
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g7
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

Query Match 100.0%; Score 3080; DB 2; Length 592;
Best Local Similarity 100.0%; Pred. No. 4.8e-223;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592


RESULT 2
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain MC
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
  Accession: B81009
  Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

  Query Match             98.2%;  Score 3024;  DB 2;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 7.6e-219;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 3
D91241
hypothetical protein ECs4900 [imported] - Escherichia coli (strain O157:H7, substrai
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: D91241
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han,
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D91241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38323.1; PID:g13364376; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4900
C;Superfamily: birA bifunctional protein

  Query Match              9.9%;  Score 305.5;  DB 2;  Length 321;
  Best Local Similarity   31.0%;  Pred. No. 2.8e-15;
  Matches  103;  Conservative   51;  Mismatches  145;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |   :|        ||: |   |          ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
                |  |: : :|| :  ||:  |    :     |:|  :|:  ||   |:    ||   |
Db         59 -SLPEPIQLLNAEQI--LGQLDGGSVTVLPVIDSTNQYLLD--RIGELKSGDA-CVAEYQ 112

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQI 176
                ||||:||||    |  |  |  |  :: |   :| || |  :     | :||  | ::
Db        113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              |||||| :  || |||:| | :||     |:| |||  : : || :   |
Db        172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy        236 GNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 4
A86089
hypothetical protein birA [imported] - Escherichia coli (strain O157:H7, substrain E
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A86089
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86089
A;Status: preliminary
A;Molecule type: DNA
```

```
A;Residues: 1-321 <STO>
A;Cross-references: GB:AE005174; NID:g12518899; PIDN:AAG59173.1; GSPDB:GN00145; UWGP:Z55
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: birA
C;Superfamily: birA bifunctional protein

  Query Match             9.9%;  Score 305.5;  DB 2;  Length 321;
  Best Local Similarity   31.0%;  Pred. No. 2.8e-15;
  Matches  103;  Conservative   51;  Mismatches  145;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |    :|        ||: |    |           ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
               |  |: : :|| : ||:  |     :     |:|  :|:  ||   |:    ||   |
Db         59 -SLPEPIQLLNAEQI--LGQLDGGSVTVLPVIDSTNQYLLD--RIGELKSGDA-CVAEYQ 112

 y         119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQI 176
               ||||:||||    |  |  |  |  :: |   :| || |  :     | :||  | ::
Db         113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy         177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
               |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db         172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy         236 GNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db         231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy         296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
               :|:| || | ||     :  : |||||| ::
Db         290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 5
BVECBF
biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Escherichia coli (strain K-12)
N;Contains: biotin operon repressor; biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 31-Dec-1988 #text_change 03-Jun-2002
C;Accession: B24029; B45727; D65204; A24989
R;Howard, P.K.; Shaw, J.; Otsuka, A.J.
Gene 35, 321-331, 1985
A;Title: Nucleotide sequence of the birA gene encoding the biotin operon repressor and b
A;Reference number: A91525; MUID:86006282; PMID:3899863
A;Accession: B24029
A;Molecule type: DNA
A;Residues: 1-321 <HOW>
A;Cross-references: GB:M10123; NID:g145430; PIDN:AAA23520.1; PID:g145432
R;Buoncristiani, M.R.; Howard, P.K.; Otsuka, A.J.
Gene 44, 255-261, 1986
A;Title: DNA-binding and enzymatic domains of the bifunctional biotin operon repressor (
A;Reference number: A24989; MUID:87055242; PMID:3536662
A;Contents: annotation; domain information
A;Note: this paper reports mutational studies done to understand DNA-binding and enzymat
R;Song, W.J.; Jackowski, S.
J. Bacteriol. 174, 6411-6417, 1992
A;Title: Cloning, sequencing, and expression of the pantothenate kinase (coaA) gene of E
A;Reference number: A45727; MUID:93015690; PMID:1328157
A;Accession: B45727
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 226-321 <SON>
A;Cross-references: GB:M90071
A;Note: residues 246-247 are correct to paper
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D65204
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-321 <BLAT>
A;Cross-references: GB:AE000471; GB:U00096; NID:g1790404; PIDN:AAC76951.1; PID:g179
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: birA
A;Map position: 89 min
C;Function: <ENZ>
A;Description: EC 6.3.4.15 [validated; MUID:99400973]; biotin-[acetyl-CoA-carboxyla
otin-accepting proteins
C;Function: <REP>
A;Description: acts as a biotin operon repressor [validated; MUID:99400973]; the re
orepressor
A;Note: the switch of birA from a biotin ligase to a repressor reflects competition
C;Superfamily: birA bifunctional protein
C;Keywords: ATP; biotin metabolism; DNA binding; ligase; repressor; transcription re
F;1-26/Domain: DNA binding #status predicted <DNA>
F;83-119,189-235/Domain: biotin activation #status predicted <BIO>

  Query Match             9.9%;  Score 304.5;  DB 1;  Length 321;
  Best Local Similarity   30.7%;  Pred. No. 3.3e-15;
  Matches  102;  Conservative   53;  Mismatches  144;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY 58
              :::| ||:|      ||     |    :|        ||: |    |           ||
Db          9 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY 58

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
               |  |: : :|: :  ||:  |    |:      |:|  :|:  ||   |:    |:   |
Db         59 -SLPEPIQLLNAKQI--LGQLDGGSVAVLPVIDSTNQYLLD--RIGELKSGDA-CIAEYQ 112

Qy         119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQI 176
               ||||:||||    |  |  |  |  :: |   :| || |  :     | :||  | ::
Db         113 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 171

Qy         177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
               |||||| :   || |||:| | :||     |:| |||  : : || :   |
Db         172 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 230

Qy         236 GNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db         231 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 289

Qy         296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
               :|:| || | ||     :  : |||||| ::
Db         290 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 321


RESULT 6
E82337
birA bifunctional protein VC0319 [imported] - Vibrio cholerae (strain N16961 serogrc
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: E82337
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Seller
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: E82337
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-320 <HEI>
A;Cross-references: GB:AE004120; GB:AE003852; NID:g9654727; PIDN:AAF93492.1; GSPDB:G
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0319
A;Map position: 1
C;Superfamily: birA bifunctional protein

  Query Match             9.3%;  Score 287.5;  DB 2;  Length 320;
```

```
  Best Local Similarity   28.6%;  Pred. No. 6.2e-14;
  Matches   97;  Conservative   45;  Mismatches  146;  Indels   51;  Gaps   11;

Qy         11 VLAELADGLPQHVSQ-------LARMADMKPQQLNGFWQQMPAHIRGLLR--------QH 55
              :| :||||   |  :       ::| |  |              ||:|:          |
Db         10 ILKQLADG-DFHSGEVLGAQLGISRAAISK------------HIQGIRDWGVDVFRVQG 55

Qy         56 DGYWRLVRPLAVFDAEGLRELGERSGFQTALKHE------CASSNDEILELARIAPDKAH 109
               || :| : : : |          :|  |: : :          |:|  :|:
Db         56 KGY-QLAQAMTLLD---------QSVIQSQVNNPVELHPIIGSTNQYLLDHVETL---VS 102

Qy        110 KTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSR 169
               |:|:   |: ||||:|| |    |  |  |  |  |         || |  ||   ||
Db        103 GTVCLAEYQASGRGRRGRHWVSPFGANLYLSIYWRLDAGMAAAMGLSLVVGVAIVEALEA 162

Qy        170 LGLK-TQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQS 227
              :|:   ::||||||     || |||:| : : |     |:|:|||  :     |
Db        163 MGVDGVKLKWPNDLYYQDKKLAGILVEMSGQAGAAAHLVIGMGINLAMRDNEGNIDQPWI 222

Qy        228 LFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDG 287
              : :   |   |  |:| ||  | ||   |  ||  :   :  |: | ||  |
Db        223 SLAEVTGQSRIDRNALAINLIAALDRTLRQYEISGMQNFVERWNRWDNFIGRPVKLLM-G 281

Qy        288 ETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLRSDD 326
                  |  :|:|  | : ||| || :: : ||||| :|
Db        282 ANEVRGIERGIDEHGGVLLETEEGLKSFIGGEISLRKND 320


RESULT 7
A83112
BirA bifunctional protein PA4280 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83112
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathc
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83112
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07668.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: birA; PA4280
C;Superfamily: birA bifunctional protein

  Query Match             9.1%;  Score 281.5;  DB 2;  Length 312;
  Best Local Similarity   33.1%;  Pred. No. 1.7e-13;
  Matches   90;  Conservative   39;  Mismatches  124;  Indels   19;  Gaps    8;

Qy         57 GYWRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTH 116
              || ||  ||:: |   :  |     :    :     |:| | :   |:|   :   : |
Db         52 GY-RLASPLSLLDPRKIDSL-----WPVEIMPTVDSTNSEAMR--RLAAGASTPFVIVAE 103

Qy        117 LQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLK-TQ 175
               |: ||||:|| |    || | :|          ||  || |  :|  |||  ||:: |
Db        104 QQTAGRGRRGRAWVSPYGENLYYSLAITVRGGARELEGLSLVVGLAVLRALESLGVQDTG 163

Qy        176 IKWPNDLVVGRDKLGGILIETVRTGGKTV----AVVGIGINFVLPKEVENAASVQSLFQT 231
              :||||||: |  |: |||:|     |        |:|||||  :    :       || :
Db        164 LKWPNDLLQGGKKIAGILLE---LSGDPADLCHVVIGIGINVNMKTATDIGQPWTSLREA 220

Qy        232 ASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVF 291
                  | : ||  |  :|   | :::|:|||    |::: |   | | ::|    :
Db        221 LGEL--VDRSRLLSALNRQLADYLDRHSREGFAASREEWESCNLWQG-AQVILSTALSSI 277

Qy        292 EGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              || | ||| :| | ||    :|    ||:|||
Db        278 EGRVLGVDSRGALRLEVDGVEQGFSGGELSLR 309


RESULT 8
AG0457
biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) [imported] - Yersinia pestis (s
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C;Accession: AG0457
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice,
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan,
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barre
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AG0457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-319 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93227.1; PID:g15981675; GSPDB:GN00175
C;Genetics:
A;Gene: birA
C;Superfamily: birA bifunctional protein
C;Keywords: ligase

  Query Match             9.0%;  Score 276.5;  DB 2;  Length 319;
  Best Local Similarity   29.0%;  Pred. No. 4.1e-13;
  Matches   96;  Conservative   46;  Mismatches  150;  Indels   39;  Gaps   12;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPLAVFD 69
              |::: |:||      ||     |    :|   |    ||         |     | ||
Db          9 RLVSILSDGFFHSGEQLGETLGMSRAAINKHMQT----IRD--------W----GLDVFT 52

Qy         70 AEG--------LRELGER-------SGFQTALKHECASSNDEILELARIAPDKAHKTICV 114
                |        :: | |:       :|  | |     |:|  :|:  ||   |:   ||
Db         53 VPGKGYSLPTSIQLLDEQKILSYLPTGHVTVLP-VVDSTNQYLLD--RITELKSGDA-CV 108

Qy        115 THLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-K 173
               |  ||||:||:|    |  |  |  |  ::       || |  :     |  ||  |
Db        109 AEYQQAGRGRRGRQWVSPFGANLYLSMYWRLEQGPAAAMGLSLVVGIVMAEVLQNLGAEK 168

Qy        174 TQIKWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTA 232
              ::||||||  :   || |||:| | :||     |:| |||  :     || : |
Db        169 VRVKWPNDLYLNDKKLAGILVELTGKTGDAAQLVIGAGINLTMRGSETNAIN-QDWINLQ 227

Qy        233 SRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFE 292
                |   |   :|:||   ::::  :| : |:: :|  :    : | |:   : :|
Db        228 DAGVTIDRNKLTAEILSELRLAVVKFENEGLSAFLSRWQKMDNYLNRPVKLIIGNQEIF- 286

Qy        293 GTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
              |  :|:| || | ||     :  : ||||||
Db        287 GIARGIDQQGALLLEQDGNIKPYIGGEISLR 317


RESULT 9
AG0934
bifunctional protein, biotin operon repressor and biotin-[acetyl-CoA carboxylase] sy
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG0934
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churc
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
A;Reference number: AB0502; PMID:11677608
A;Accession: AG0934
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-320 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD09497.1; PID:g16504614; GSPDB:GN00176
C;Genetics:
```

```
A;Gene: STY3741
C;Superfamily: birA bifunctional protein

  Query Match             8.8%;  Score 269.5;  DB 2;  Length 320;
  Best Local Similarity  29.6%;  Pred. No. 1.4e-12;
  Matches   97;  Conservative   50;  Mismatches  148;  Indels   33;  Gaps   13;

Qy         11 VLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GYW 59
              ::: ||||      ||      |    :|        ||: |    |          ||
Db         10 LISLLADGEFHSGEQLGERLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY- 58

Qy         60 RLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQS 119
               |  |: : ||: :    :      |:     |:|  :|:  ||   ::    ||   |
Db         59 SLPEPIQLLDADRIH--SQLDSGNVAVLPVIDSTNQYLLD--RIGELRSGDA-CVAEYQQ 113

Qy        120 KGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQIK 177
              ||||:||||    |  |  |  |  :: |   :| || |  :      |  ||  | ::|
Db        114 AGRGRRGRKWFSPFGANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRELGADKVRVK 172

Qy        178 WPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRG 236
              |||||  :   || |||:| | :||     |:| |||  : : ||     |
Db        173 WPNDLYLLDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGI 231

Qy        237 NADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVK 296
                |  :|   |: :| | |  : ::| :|::: ::  :    : | |:   : :| |  :
Db        232 TLDRNMLAAKLIYKLRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLIIGDKEIF-GISR 290

Qy        297 GVDGQGVLHLETAEGKQTVVSGEISLRS 324
              |:| || | ||     :  : |||||||
Db        291 GIDTQGALLLEQDGVIKPWMGGEISLRS 318


RESULT 10
I40327
baf protein - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40327; S70669
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
J. Bacteriol. 177, 3801-3807, 1995
A;Title: Identification of a Bordetella pertussis regulatory factor required for transcr
A;Reference number: I40327; MUID:95325323; PMID:7601846
A;Accession: I40327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229
;Allen, A.; Maskell, D.
 ol. Microbiol. 19, 37-52, 1996
A;Title: The identification, cloning and mutagenesis of a genetic locus required for lip
A;Reference number: S70669; MUID:96419162; PMID:8821935
A;Accession: S70669
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 239-267 <ALL>
A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: baf

  Query Match             8.3%;  Score 256.5;  DB 2;  Length 267;
  Best Local Similarity  32.0%;  Pred. No. 1e-11;
  Matches   85;  Conservative   43;  Mismatches  113;  Indels   25;  Gaps    9;

Qy        341 LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKVDGNVRIVGCAVCG 395
              :|:| ||||||  | : :   |    ||      ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        396 EFKKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 449
                :   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        450 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTG 509
                :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        510 NAVASGMMDAVCGSVM--MMHGRLKEKTGAGKPVDVIITGGG-------AAKVAEALPPA 560
                |:|||:  |  |:::   : || :     |:  :: : |||       | ::      |
Db        183 QAIASGIAAAQAGAIVRQWLAGRQR----YGQAPEIYVAGGGWPEVRQEAERLLAVTGAA 238

Qy        561 FLAENTVRVADNLVIHGLLNLIAAEG 586
              | |       |: |: ||  | ||:|
Db        239 FGATPQPTYLDSPVLDGLAAL-AAQG 263


RESULT 11
D75285
birA bifunctional protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: D75285
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: D75285
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-474 <WHI>
A;Cross-references: GB:AE002065; GB:AE000513; NID:g6460149; PIDN:AAF11888.1; PID:g64
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2341
A;Map position: 1

  Query Match             7.9%;  Score 244;  DB 2;  Length 474;
  Best Local Similarity  32.5%;  Pred. No. 2e-10;
  Matches  111;  Conservative   41;  Mismatches  122;  Indels   68;  Gaps   20;

Qy          9 WRVLAELADGL---PQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWRLVRPL 65
              ||: | |   |   ||    | |  |:    :|   :::         | ||      |:
Db        167 WRMPARLLPLLTDRPQSGDDLGRALDLGRVTVNTLARRL---------QEDGV-----PV 212

Qy         66 AVFDAEGLRELGE--------RSGFQTALKHE--CASSNDEILELARIAPDK-----AHK 110
               :  |    | |         |  |  ||:::    |: ||:    |   |       |
Db        213 LISRAGYALEPGTPAPGLVPLRGQFGRALRYQGTVTSTQDEV----RFWADSERGPAPHG 268

Qy        111 TICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSL--SPVAA-VACRRAL 167
               : |   |: ||||:|| |    |  | ||   |  |    |  |   |:|| ||   |
Db        269 AVVVAERQTAGRGRRGRVWDTTHG-TLAFS---VLLREALSLADLVRMPLAAGVALHAAC 324

Qy        168 SRLGLKTQIKWPNDLVV--GRDKLGGILIETVRTGGKT-VAVVGIGINFVLPKEVENAAS 224
              : ||:   :||||||:   || || |||:|    | :   ||:|||||  :    | ||
Db        325 TALGVPCGLKWPNDLLTPDGR-KLAGILLEADLRGEEARRAVLGIGIN--VGTAPEGAAH 381

Qy        225 VQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLL 284
              :       :|     |||| | |  :|:  | | |       :| ::| :   |: | :
Db        382 LHESVPDLTR-----AAVLGEVLF-QLEHWLAQPA----PAIIATWKATSLTLGRPVRV- 430

Qy        285 RDGETVFEGTVKGV----DGQGVLHLETAEGKQTVVSGEISL 322
                  |   |||:||    | || | :||| | :|| :|:: |
Db        431 ----TTARGTVEGVARDLDEQGNLLVETAAGLETVSAGDVQL 468


RESULT 12
B82637
bifunctional biotin operon/biotin acetyl-CoA-carboxylase synthetase transcription re
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C;Accession: B82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
```

```
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: B82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-317 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84604.1; GSPDB:GN001
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigr
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
C;Contents: annotation
C;Genetics:
A;Gene: XF1796
C;Superfamily: birA bifunctional protein

  Query Match             7.5%;  Score 232.5;  DB 2;  Length 317;
  Best Local Similarity   28.3%;  Pred. No. 8.2e-10;
  Matches   82;  Conservative   46;  Mismatches  121;  Indels   41;  Gaps   11;

Qy          53 RQHDGYWRLVRPLAVFDAEGLRELGERSGFQTALKHECA------------SSNDEILE 99
               |   ||  |  ||:: :|| :|           | |  |              |:|  :|
Db          48 RAEQGY-ALAAPLSLLNAERIR---------AELSHSIAAELRVLDVAWTLESTNTSLLA 97

Qy         100 LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVA 159
                 :|      : :  |: |||::|| |   |   :  |   ||     :|  ||
Db          98 RPALADG---VEVLLAERQTGGRGQRGRIWISPLAANIYLSLSRVFSSGAAQLMGLSVAV 154

Qy         160 AVACRRALSRLGLK-TQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLP-- 216
                ||   ||   |    ::|||||||    ||||:|:|     |:  ||:|:|:|  :|
Db         155 GVAIAEALIAAGFSHVRLKWPNDLVDNDRKLGGVLVE---RQGEARAVIGLGLNVRMPTN 211

Qy         217 --KEVENA-ASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAA 273
                 ::::     :::|  :   | ||  |:||  ||  |    |    :| |||:  |
Db         212 FAEQIDQPWVDLETLAGSTVER-NALVALLLSALLPALHLFQL----EGLAPFLPRYAKL 266

Qy         274 NRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVSGEISLR 323
               :  | ||  : :    : |  :|:   | | :|     :   :|::|:|
Db         267 DCLVGCAV-WIEEANLKYVGIARGLAADGALLVEINGVLRPFHTGQVSVR 315


RESULT 13
H83111
hypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

  Query Match             7.5%;  Score 231.5;  DB 2;  Length 248;
  Best Local Similarity   27.7%;  Pred. No. 6.9e-10;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy         341 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 397
               | ||  ||| :||  :|    :  |      |   :  | :: |      : :  | :
Db           3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56

Qy         398 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 450
                 :  :  : :||  :  | ||        |  |:|| |   :  | |||   : :   :: |
Db          57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116

Qy         451 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 503
               |:|:  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db         117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176

Qy         504 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 558
                 | | |  :     | ::|: | ::|:        |   :: :||| |  | : |
Db         177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228

Qy         559 PAFLAENTVRVADNLVIHGL 578
                |       |:  :||  ||
Db         229 GA-------RIMPDLVFVGL 241


RESULT 14
A69595
transcription repressor of biotin operon birA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C;Accession: A69595
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: A69595
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-325 <KUN>
A;Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14160.1; PID:g2634
A;Experimental source: strain 168
C;Genetics:
A;Gene: birA
C;Superfamily: Bacillus subtilis transcription repressor of biotin operon birA

  Query Match             7.4%;  Score 226.5;  DB 2;  Length 325;
  Best Local Similarity   28.0%;  Pred. No. 2.4e-09;
  Matches   86;  Conservative   50;  Mismatches  118;  Indels   53;  Gaps   13;

Qy          47 HIRGLLRQ--------HDGYWRLVRPLAVFDAE---GLRE--LGERSGFQTALK------ 87
               ||  | ::          ||  : :|  : ::|   ||:   :|:   :   |
Db          40 HIEELRKEGYEVEAVRRKGYRLIKKPGKLSESEIRFGLKTEVMGQHLIYHDVLSSTQKTA 99

Qy          88 HECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDR 147
               || |::|          ||:    |: |   |: |||  | |  : |  :  |      |
Db         100 HELANNN---------APE---GTLVVADKQTAGRGRMSRVWHSQEGNGVWMSLILRPDI 147

Qy         148 PQYELGSLSPVAAVACRRALSR-LGLKTQIKWPNDLVVGRDKLGGILIE------TVRTG 200
               |  :   |: :|||  : :    |::| ||||||:::   |  ||| |       ||:
Db         148 PLQKTPQLTLLAAVAVVQGIEEAAGIQTDIKWPNDILINGKKTVGILTEMQAEEDRVRS- 206

Qy         201 GKTVAVVGIGINF-----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVL 255
                     ::|||||       | |::: |:  || | |  :   | | ::: :|   :
```

```
Db        207 ----VIIGIGINVNQQPNDFPDELKDIAT--SLSQAAGEK--IDRAGVIQHILLCFEKRY 258

Qy        256 LQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTV 315
              |   || |    :::     |   :  |     | |   |:| :||| ||| || : :
Db        259 RDYMTHGFTPIKLLWESYALGIG-TNMRARTLNGTFYGKALGIDDEGVLLLETNEGIKKI 317

Qy        316 VSGEISL 322
               | :| |
Db        318 YSADIEL 324


RESULT 15
E87000
biotin apo-protein ligase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 17-May-2002
C;Accession: E87000
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 am, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: E87000
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-274 <STO>
A;Cross-references: GB:AL450380; NID:g13092858; PIDN:CAC30241.1; GSPDB:GN00147
C;Genetics:
A;Gene: birA
C;Superfamily: Bacillus subtilis transcription repressor of biotin operon birA

  Query Match             7.1%;  Score 218.5;  DB 2;  Length 274;
  Best Local Similarity   34.5%;  Pred. No. 7.5e-09;
  Matches   86;  Conservative   32;  Mismatches  102;  Indels   29;  Gaps   12;

Qy         89 ECASSNDEILELARIAPDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFG-WVFDR 147
              :  |:| ::|. |    |     :   | |: |||| || ||      :  | |  : |
Db         34 QTGSTNADLLARAASGIDIDGAVLIAEH-QTAGRGRHGRGWSASPRTQITMSIGVSIADV 92

Qy        148 PQYELGSLSPVAAVACRRALSRL----GLKTQIKWPNDLVVG----RDKLGGILIETVRT 199
              |      ||    ||    :: |    |:   :|||||:: |    : || ||| | ||
Db         93 PVATWAWLSLATGVAVLDTVAPLIDVTGVDAGLKWPNDVLAGPPGSKGKLAGILAEVVR- 151

Qy        200 GGKTVAVVGIGINFV-LPKEVE--NAASVQSLFQTASRRGNADAAVLLETLLAELDAVLL 256
                | |||:|:|    ||||:   | |:  |   |      |   |:  || |||| :
Db        152 ---PVIVVGVGLNVTQAPKEVDGPGATSLLDLGVPA-----PDREQLVCWLLRELDARVG 203

Qy        257 QYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK---Q 313
              |: |      :|:|:| :   |  |     |     |||:|:| || | |:| ||    |
Db        204 QW-RSADPQLMADYRARSPIIGSRVRAELPGGKDVVGTVRGIDDQGRLCLKT-EGPTKGQ 261

Qy        314 TVV--SGEI 320
              |||  :|::
Db        262 TVVISAGDV 270


Search completed: June 24, 2003, 22:03:20
Job time : 22.5927 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 60.7795 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-22
Perfect score: 3080
Sequence: 1 MTVLKPSHWRVLAELADGLP.....LVIHGLLNLIAAEGGESEHT 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3080	100.0	592	20	AAY38616	Neisseria meningit
2	3080	100.0	592	21	AAY74913	Neisseria meningit
3	3080	100.0	592	23	AAU91166	Pantothenate kinas
4	3024	98.2	592	20	AAY38615	Neisseria meningit
5	3024	98.2	592	21	AAY74912	Neisseria meningit
6	3024	98.2	592	23	AAU91169	Pantothenate kinas
7	2921	94.8	592	20	AAY38618	Neisseria gonorrho
8	2921	94.8	592	21	AAY74911	Neisseria gonorrhe
9	2346	76.2	455	21	AAY74910	Neisseria meningit
10	2223	72.2	460	23	AAU91167	Pantothenate kinas
11	2197	71.3	455	20	AAY38617	Neisseria gonorrho
12	2197	71.3	455	21	AAY74908	Neisseria gonorrhe
13	1852.5	60.1	389	21	AAY74909	Neisseria meningit
14	969.5	31.5	189	20	AAY38614	Neisseria meningit
15	304.5	9.9	322	20	AAY31601	Escherichia coli b
16	269	8.7	363	22	ABG28654	Novel human diagno
17	256.5	8.3	267	23	AAU91162	Bordella pertussis
18	256	8.3	302	22	AAB88546	Haemophilus influe
19	256	8.3	302	23	AAU91477	Haemophilus influe
20	243	7.9	249	23	AAU91178	Pantothenate kinas
21	231.5	7.5	248	23	AAU91164	Pantothenate kinas
22	217.5	7.1	325	23	ABB48003	Listeria monocytog
23	210	6.8	291	22	AAU04901	Micromonospora eve
24	201.5	6.5	282	22	AAU64093	Propionibacterium
25	198.5	6.4	242	23	AAU91180	Pantothenate kinas
26	190	6.2	294	22	ABB09064	Thermus caldophilu
27	189.5	6.2	237	22	AAB96079	Putative P. abyssi
28	187	6.1	249	23	AAU91182	Pantothenate kinas
29	183.5	6.0	269	22	AAB76593	Corynebacterium gl
30	183.5	6.0	288	22	AAG90532	C glutamicum prote
31	182.5	5.9	323	21	AAB07758	Amino acid sequenc
32	180.5	5.9	265	23	AAU91151	Streptomyces coeli
33	179	5.8	311	21	AAY70846	Streptococcus pneu
34	175	5.7	260	23	AAU91173	Pantothenate kinas
35	173.5	5.6	323	22	ABB09053	Thermus caldophilu
36	169	5.5	323	23	ABB55129	Lactococcus lactis
37	168.5	5.5	289	23	ABP30290	Streptococcus poly
38	168.5	5.5	311	23	ABP27627	Streptococcus poly
39	168	5.5	250	23	AAU91150	Clostridium acetob
40	168	5.5	262	23	AAU91170	Pantothenate kinas
41	166	5.4	250	23	ABB55254	Lactococcus lactis
42	161	5.2	254	21	AAG32612	Arabidopsis thalia
43	161	5.2	266	21	AAG06098	Arabidopsis thalia
44	161	5.2	266	21	AAG49376	Arabidopsis thalia
45	161	5.2	317	21	AAG32611	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAY38616
ID AAY38616 standard; Protein; 592 AA.
XX
AC AAY38616;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis strain A antigen encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX

```
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match            100.0%;  Score 3080;  DB 20;  Length 592;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-279;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592


RESULT 2
AAY74913
ID   AAY74913 standard; Protein; 592 AA.
XX
AC   AAY74913;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1300.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53675.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 715; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match            100.0%;  Score 3080;  DB 21;  Length 592;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-279;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
```

```
Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
```

RESULT 3
AAU91166
ID AAU91166 standard; Protein; 592 AA.
XX
AC AAU91166;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #4.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria meningitidis.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 85-86; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 592 AA;

```
  Query Match           100.0%;  Score 3080;  DB 23;  Length 592;
  Best Local Similarity  100.0%;  Pred. No. 2.5e-279;
  Matches  592;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
```

RESULT 4
AAY38615
ID AAY38615 standard; Protein; 592 AA.
XX
AC AAY38615;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.

```
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
     organisms are closely related. Fragments of the nucleic acids
     are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             98.2%;  Score 3024;  DB 20;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 4.4e-274;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              ||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 5
AAY74912
ID   AAY74912 standard; Protein; 592 AA.
XX
AC   AAY74912;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53674.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 714; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             98.2%;  Score 3024;  DB 21;  Length 592;
  Best Local Similarity   98.5%;  Pred. No. 4.4e-274;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
```

```
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
```

RESULT 6
AAU91169
ID AAU91169 standard; Protein; 592 AA.
XX
AC AAU91169;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #7.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria meningitidis.
XX
PN WO200216601-A2.
XX
D 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54190.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein
PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 6; Page 96-98; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 592 AA;

Query Match 98.2%; Score 3024; DB 23; Length 592;
Best Local Similarity 98.5%; Pred. No. 4.4e-274;
Matches 582; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

```
Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
```

RESULT 7
AAY38618
ID AAY38618 standard; Protein; 592 AA.
XX
AC AAY38618;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigenic protein encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.

```
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12075.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.

CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             94.8%;  Score 2921;  DB 20;  Length 592;
  Best Local Similarity   94.9%;  Pred. No. 1.9e-264;
  Matches  561;  Conservative   12;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||| ||||||||||||||:||||||  |||:|||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||||||  ||| ||| :|||||: ||: ||||||||||||||||| |||||||||
Db        241 AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              :|||||||||:|||||||||| |:| |||||| ||||||||:|||||||||||||||||
Db        301 RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||| |||||||||||| |||||:|||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              |||||||||||||||| ||||||||||||||||||||||||||:|||||||||| |||||
Db        481 HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 8
AAY74911
ID   AAY74911 standard; Protein; 592 AA.
XX
AC   AAY74911;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1296.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53673.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 713; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             94.8%;  Score 2921;  DB 21;  Length 592;
  Best Local Similarity   94.9%;  Pred. No. 1.9e-264;
  Matches  561;  Conservative   12;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLAREADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
```

```
Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRDLGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180
              |||||||||||||||||||||| |||||||||||||||:||||||  |||:|||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||||| ||| ||| :|||||: ||: |||||||||||||||| ||||||||||
Db        241 AVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              :|||||||||:|||||||||| |:| |||||| ||||||||:||||||||||||||||
Db        301 RGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||| |||||||||||||| ||||||:|||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              |||||||||||||||| |||||||||||||||||||||||:|||||||||| |||||
Db        481 HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 9
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             76.2%;  Score 2346;  DB 21;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-211;
  Matches  455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 120

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 180

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        378 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        438 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 360

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 420

Qy        558 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 592
              ||||||||||||||||||||||||||||||||||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHT 455


RESULT 10
AAU91167
ID   AAU91167 standard; Protein; 460 AA.
XX
AC   AAU91167;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #5.
```

```
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
     N-PSDB; ABK54188.

PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 8; Page 92-94; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   460 AA;

  Query Match             72.2%;  Score 2223;  DB 23;  Length 460;
  Best Local Similarity   93.7%;  Pred. No. 3.2e-199;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||||
Db          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60

Qy        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120

Qy        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||||:
Db        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180

Qy        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||||
Db        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240

Qy        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432
              |||||||| |||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300

Qy        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360

Qy        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552
              |||| |||||||||||||||||||||||:|||||||| ||||||||||||||||||||
Db        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420

Qy        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||||||||
Db        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459


RESULT 11
AAY38617
ID   AAY38617 standard; Protein; 455 AA.
XX
AC   AAY38617;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   455 AA;

  Query Match             71.3%;  Score 2197;  DB 20;  Length 455;
  Best Local Similarity   93.8%;  Pred. No. 8.6e-197;
  Matches  426;  Conservative   11;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 197
              |||||| ||||||||||||||:|||||  |||:|||||||||||||||||||||||||
Db          1 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 257
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |
Db         61 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 120

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
              || :|||||: ||: |||||||||||||||||| |||||||||:||||||||:|||||
Db        121 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 180

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
```

```
                ||||||  |:|  ||||||  ||||||||:|||||||||||||||||||||||||||||||||
Db        181 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        378 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 437
                |||| ||||||||||||||  |||||:|||||||||||||||||||||||||||||||||||
Db        241 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        438 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 360

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
                |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 420

Qy        558 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
                |||||||||||||||||||||||||||||||||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 454
```

RESULT 12
AAY74908
ID AAY74908 standard; Protein; 455 AA.
XX
AC AAY74908;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1290.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrheae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
DR WPI; 2000-062150/05.
DR N-PSDB; AAZ53670.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 709; 1453pp; English.
XX
CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 455 AA;

```
  Query Match             71.3%;  Score 2197;  DB 21;  Length 455;
  Best Local Similarity   93.8%;  Pred. No. 8.6e-197;
  Matches  426;  Conservative   11;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 197
                ||||||  |||||||||||||||:||||||  |||:||||||||||||||||||||||||
Db          1 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 257
                | |||||||||||||||||||||||||||||||||||||||||||||||||| ||| |
Db         61 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 120

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317
                || :|||||: ||: |||||||||||||||||| |||||||||:||||||||||:|||||
Db        121 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 180

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
                ||||||  |:|  ||||||  ||||||||:|||||||||||||||||||||||||||||||
Db        181 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        378 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 437
                ||||  ||||||||||||||  |||||:|||||||||||||||||||||||||||||||||||
Db        241 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        438 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 360

Qy        498 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 557
                |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 420

Qy        558 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
                |||||||||||||||||||||||||||||||||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 454
```

RESULT 13
AAY74909
ID AAY74909 standard; Protein; 389 AA.
XX
AC AAY74909;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.

```
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53671.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 710; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   389 AA;

  Query Match             60.1%;  Score 1852.5;  DB 21;  Length 389;
  Best Local Similarity   81.3%;  Pred. No. 1.3e-164;
  Matches  370;  Conservative    2;  Mismatches   16;  Indels   67;  Gaps    2;

Qy        138 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 197
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||
Db          1 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 60

Qy        198 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 257
              |||||||||||||||||| ||||||||||||||||||||||
Db         61 RTGGKTVAVVGIGINFVLPXEVENAASVQSLFQTASRRGNADAAVLLXX---------- 109

Qy        258 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 317

Qy        110 ----------------------------------------------------XXXXX 114

Qy        318 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 377
              ||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        115 XEISLRSDXRPVSVXKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 174

Qy        378 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL-GIRNHYRHPEEHGSDR 436
              |||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        175 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALFGIRNHYRHPEEHGSDR 234

Qy        437 WFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 496
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 WFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 294

Qy        497 HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEA 556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEA 354

Qy        557 LPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591
              |||||||||||||||||:||||:||||| | ||
Db        355 LPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 389

RESULT 14
```

```
AAY38614
ID   AAY38614 standard; Protein; 189 AA.
XX
AC   AAY38614;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12072.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 171; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   189 AA;

  Query Match             31.5%;  Score 969.5;  DB 20;  Length 189;
  Best Local Similarity   97.4%;  Pred. No. 2.1e-82;
  Matches  184;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy        319 EISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEW 378
              ||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 EISLRSDXRPVSVXKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEW 60

Qy        379 AEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWF 438
              ||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db         61 AEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAXGIRNHYRHPEEHGSDRWF 120

Qy        439 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLG-GTIMPGFHLMKESLAVRTANLNRH 497
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        121 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGXGTIMPGFHLMKESLAVRTANLNRH 180

Qy        498 AGKRYPFPT 506
              |||||||||
Db        181 AGKRYPFPT 189

RESULT 15
```

```
AAY31601
ID   AAY31601 standard; Protein; 322 AA.
XX
AC   AAY31601;
XX
DT   09-NOV-1999  (first entry)
XX
DE   Escherichia coli biotin synthesis repressor enzyme BirA.
XX
KW   BirA; DNA primer; PCR; polymerase chain reaction;
KW   biotin synthesis repressor; biotin.
XX
OS   Escherichia coli.
XX
PN   WO9937785-A1.
XX
PD   29-JUL-1999.
XX
PF   25-JAN-1999;   99WO-GB00245.
XX
PR   26-JAN-1998;   98GB-0001621.
XX
PA   (ISIS-) ISIS INNOVATION LTD.
XX
PI   Jakobsen BK,  O'Callaghan CA;
XX
DR   WPI; 1999-469137/39.
DR   N-PSDB; AAZ08957.
XX
PT   Production of a bacterial repressor (BirA enzyme), useful for the
PT   specific labeling of proteins with biotin
XX
PS   Claim 6; Fig 1; 14pp; English.
XX
CC   The present sequence is an Escherichia coli BirA enzyme which is a
CC   biotin biosynthesis repressor.  DNA primers were used to amplify the BirA
CC   gene using DNA derived from E. coli, for subsequent cloning of the gene
CC   into plasmid pGEX-2T, containing a thrombin cleavage recognition sequence
CC   between a sequence encoding a glutathione-S-transferase tag and the
CC   cloning site into which the BirA cloning sequence was inserted.  The
CC   resulting plasmid, encoding BirA as a fusion protein, was designated
CC   COC 053.  The plasmid was then used to transform E. coli
CC   BL21(DE3) pLysS(4) for recombinant BirA enzyme production.  The method
CC   is used for specific transfer of a biotin group to a protein containing
CC   a specific target sequence, in order to specifically label an engineered
CC   protein with biotin.  The method also produces higher quality amounts
CC   of the BirA enzyme compared to existing methods.
XX
SQ   Sequence   322 AA;

  Query Match             9.9%;  Score 304.5;  DB 20;  Length 322;
  Best Local Similarity   30.7%;  Pred. No. 9.6e-20;
  Matches  102;  Conservative   53;  Mismatches  144;  Indels   33;  Gaps   13;

Qy         10 RVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHD----------GY  58
              :::| ||:|      ||     |    :|        ||: |    |          ||
Db         10 KLIALLANGEFHSGEQLGETLGMSRAAIN-------KHIQTL---RDWGVDVFTVPGKGY  59

Qy         59 WRLVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQ 118
               |  |: : :|: :  ||:  |   |:     |:|  :|:  ||   |:    |:   |
Db         60 -SLPEPIQLLNAKQI--LGQLDGGSVAVLPVIDSTNQYLLD--RIGELKSGDA-CIAEYQ 113

Qy        119 SKGRGRQGRKWSHRLGECLMFSFGWVFDR-PQYELGSLSPVAAVACRRALSRLGL-KTQI 176
              ||||:||||    |  |  |  |  :: |   :| || |  :     | :||  | ::
Db        114 QAGRGRRGRKWFSPFGANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRV 172

Qy        177 KWPNDLVVGRDKLGGILIE-TVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRR 235
              ||||||  :   || |||:| | :||     |:| |||  : : || :   |
Db        173 KWPNDLYLQDRKLAGILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAG 231

Qy        236 GNADAAVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTV 295
               | |   |   |: || | |  : ::| ||::: ::  :    : | |:   : :| |
Db        232 INLDRNTLAAMLIRELRAALELFEQEGLAPYLSRWEKLDNFINRPVKLIIGDKEIF-GIS 290

Qy        296 KGVDGQGVLHLETAEGKQTVVSGEISLRSDDR 327
              :|:| || | ||     :  : |||||| ::
Db        291 RGIDKQGALLLEQDGIIKPWMGGEISLRSAEK 322


Search completed: June 24, 2003, 21:46:27
Job time : 62.7795 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 26.9909 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-43
Perfect score: 3078
Sequence: 1 MTVLKLSHWRVLAELADGLP.....LVIYGLLNMIAAEGREYEHI 592

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3078	100.0	592	9	US-09-813-453A-43	Sequence 43, Appl
2	3024	98.2	592	9	US-09-813-453A-22	Sequence 22, Appl
3	2197	71.4	460	9	US-09-813-453A-39	Sequence 39, Appl
4	259	8.4	302	9	US-10-260-877-110	Sequence 110, App
5	254.5	8.3	267	9	US-09-813-453A-15	Sequence 15, Appl
6	245	8.0	249	9	US-09-813-453A-61	Sequence 61, Appl
7	234.5	7.6	248	9	US-09-813-453A-20	Sequence 20, Appl
8	200	6.5	242	9	US-09-813-453A-65	Sequence 65, Appl
9	194	6.3	311	10	US-09-859-270-2	Sequence 2, Appli
10	191	6.2	249	9	US-09-813-453A-70	Sequence 70, Appl
11	184.5	6.0	288	9	US-09-738-626-4286	Sequence 4286, Ap
12	177	5.8	265	9	US-09-813-453A-4	Sequence 4, Appli
13	172	5.6	260	9	US-09-813-453A-51	Sequence 51, Appl
14	169.5	5.5	262	9	US-09-813-453A-45	Sequence 45, Appl
15	168	5.5	250	9	US-09-813-453A-3	Sequence 3, Appli
16	162.5	5.3	257	9	US-09-813-453A-53	Sequence 53, Appl
17	161	5.2	244	9	US-09-813-453A-41	Sequence 41, Appl
18	150	4.9	219	9	US-09-813-453A-57	Sequence 57, Appl
19	150	4.9	258	9	US-09-813-453A-2	Sequence 2, Appli
20	147.5	4.8	272	9	US-09-712-363-276	Sequence 276, App
21	147.5	4.8	272	9	US-09-813-453A-5	Sequence 5, Appli
22	147.5	4.8	273	9	US-09-813-453A-10	Sequence 10, Appl
23	146	4.7	246	9	US-09-813-453A-9	Sequence 9, Appli
24	146	4.7	258	9	US-09-813-453A-49	Sequence 49, Appl
25	141	4.6	257	9	US-09-813-453A-13	Sequence 13, Appl
26	140.5	4.6	256	9	US-09-813-453A-55	Sequence 55, Appl
27	139.5	4.5	255	9	US-09-813-453A-7	Sequence 7, Appli
28	136.5	4.4	212	9	US-09-813-453A-59	Sequence 59, Appl
29	136	4.4	254	9	US-09-813-453A-47	Sequence 47, Appl
30	134	4.4	258	9	US-09-813-453A-6	Sequence 6, Appli
31	133	4.3	233	9	US-09-813-453A-17	Sequence 17, Appl
32	116.5	3.8	223	9	US-09-895-913A-74	Sequence 74, Appl
33	116.5	3.8	223	9	US-09-813-453A-14	Sequence 14, Appl
34	116.5	3.8	223	9	US-09-813-453A-67	Sequence 67, Appl
35	115.5	3.8	241	9	US-09-813-453A-63	Sequence 63, Appl
36	114.5	3.7	3816	9	US-09-808-880-3	Sequence 3, Appli
37	105.5	3.4	4613	9	US-09-860-846-31	Sequence 31, Appl
38	105.5	3.4	4613	9	US-09-988-384B-31	Sequence 31, Appl
39	105.5	3.4	4613	9	US-09-836-821-31	Sequence 31, Appl
40	105.5	3.4	4613	10	US-09-861-289-31	Sequence 31, Appl
41	105.5	3.4	11877	9	US-09-860-846-6	Sequence 6, Appli
42	105.5	3.4	11877	9	US-09-836-821-6	Sequence 6, Appli
43	105.5	3.4	11877	10	US-09-861-289-6	Sequence 6, Appli
44	105.5	3.4	12199	9	US-09-988-384B-6	Sequence 6, Appli
45	104.5	3.4	340	10	US-09-815-242-11743	Sequence 11743, A

ALIGNMENTS

RESULT 1
US-09-813-453A-43
; Sequence 43, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-813-453A-43

Query Match 100.0%; Score 3078; DB 9; Length 592;
Best Local Similarity 100.0%; Pred. No. 6.4e-257;
Matches 592; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEHI 592


RESULT 2
US-09-813-453A-22
; Sequence 22, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-22

  Query Match            98.2%;  Score 3024;  DB 9;  Length 592;
  Best Local Similarity  98.5%;  Pred. No. 2.9e-252;
  Matches  582;  Conservative    2;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 MTVLKLSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MTVLKPSHWRVLAELADGLPQHVSQLARMADMKPQQLNGFWQQMPAHIRGLLRQHDGYWR 60

Qy         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LVRPLAVFDAEGLRELGERSGFQTALKHECASSNDEILELARIAPDKAHKTICVTHLQSK 120

Qy        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPN 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
Db        121 GRGRQGRKWSHRLGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPN 180

Qy        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 DLVVGRDKLGGILIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADA 240

Qy        241 AVLLETLLVELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AVLLETLLAELDAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDG 300

Qy        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 QGVLHLETAEGKQTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTF 360

Qy        361 ATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        361 ATVGSAPYRDLSPLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAL 420

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480

Qy        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 HLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKP 540

Qy        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNLIAAEGGESEH 591
              |||||||||||||||||||||||||||||||||||:||||:||||| | ||
Db        541 VDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 3
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
;   LENGTH: 460
;   TYPE: PRT
;   ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39

  Query Match            71.4%;  Score 2197;  DB 9;  Length 460;
  Best Local Similarity  92.6%;  Pred. No. 4.3e-181;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192
              :|||||||||| ||||||||||||||||:|||||  |||: |||||||||||||||||
Db          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60

Qy        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252
              |||||| |||||||||||||||||||||||||||||||||||||||||||||| ||
Db         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120

Qy        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||:
Db        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180

Qy        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372
              |||||||||| |:| |||||| |||||||:||||||||||||||||||||||||||||||
Db        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240

Qy        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432
              |||||||||||||||||||| ||||||:||||||||||||||||||||||||||||||||
Db        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300

Qy        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360

Qy        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552
              |||| |||||||||||||||||||||||||:|||||||||| ||||||||||||||||
Db        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420

Qy        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |||||||||||||||||||||:||||:||||| | ||
Db        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459


RESULT 4
US-10-260-877-110
; Sequence 110, Application US/10260877
; Publication No. US20030021813A1
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Chovan, Linda E.
;  APPLICANT:  Hessler, Paul E.
;  APPLICANT:  Reich, Karl A.
;  TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
;  TITLE OF INVENTION:  SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  'ESSENTIAL GENES'
;  FILE REFERENCE: 6565.US.P1
;  CURRENT APPLICATION NUMBER: US/10/260,877
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: US/09/649,145
;  PRIOR FILING DATE: 2000-08-25
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 110
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: H. influenzae
US-10-260-877-110

  Query Match             8.4%;  Score 259;  DB 9;  Length 302;
  Best Local Similarity   29.4%;  Pred. No. 3.2e-14;
  Matches   92;  Conservative   56;  Mismatches  131;  Indels   34;  Gaps   13;

Qy         26 LARMADMKPQ---QLNGFWQQMPAHIR-----GLLRQHDGY-WRLVRPLAVFDAEGLREL 76
              |  :|| :|:   :|  | : :   |:     ||    ||  :||| | : : : :
Db          7 LTYLADCQPKVRSELEKFSKNLEEDIQQLREIGLDILVDGQDYRLVPMLPLLNPQQI--- 63

Qy         77 GERSGFQTALKHE--CASSNDEILE--LARIAPDKAHKTICVTHLQSKGRGRQGRKWSHR 132
                 : |  :: ::   :|:|: ||:  |:    |     :||   |: ||||:||:|
Db         64 -STALFPYSIHYQPIISSTNEWILQNILSLKKGD-----LCVAEYQTAGRGRRGRQWLSP 117

Qy        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192
                :|||| | || |:  :  || |  :|         |:||:|||||::    |||||
Db        118 FAGQIMFSFYWAFD-PKKSIEGLSLVIGLAIAEV-----LNVQVKWPNDILFDERKLGGI 171

Qy        193 LIETV-RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVEL 251
              |:|       |    |:|||||  | |: | :   |   :       : :   ||  |:  |
Db        172 LVEIANHKNGMLNLVIGIGINVSLSKQTEIS---QPYAEVCEIDPDVERQTLLPKLIQHL 228

Qy        252 DAVLLQYARDGF-APFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAE 310
                |  : ::|    |   :|: |      : :| : : |  |  :|:| :| | :
Db        229 YTRLNIFEQNGIDEEFQQAWQSYNAFSNSEINVLTE-QGVISGIEQGIDERGYLKVLCGN 287

Qy        311 GKQTVVSGEISLR 323
                |    ||:|||
Db        288 KIQMFNGGEVSLR 300


RESULT 5
US-09-813-453A-15
; Sequence 15, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Bordetella pertussis
US-09-813-453A-15

  Query Match             8.3%;  Score 254.5;  DB 9;  Length 267;
  Best Local Similarity   31.6%;  Pred. No. 6.4e-14;
  Matches   84;  Conservative   44;  Mismatches  113;  Indels   25;  Gaps    9;

Qy        341 LLLDGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 395
              :|:| |||||  | : :   |    ||      ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        396 EFKKAQVQEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 449
                :   :   |       | || :   |:|:|| ||:|:: |:|||    :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        450 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTG 509
                :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        510 NAVASGMMDAVCGSVM--MMHGRLKEKTGAGKPVDVIITGGG-------AAKVAEALPPA 560
                |:|||:  |  |:::   : || :     |:  :: : |||       | ::      |
Db        183 QAIASGIAAAQAGAIVRQWLAGRQR----YGQAPEIYVAGGGWPEVRQEAERLLAVTGAA 238

Qy        561 FLAENTVRVADNLVIYGLLNMIAAEG 586
              | |      |: |: ||   :||:|
Db        239 FGATPQPTYLDSPVLDGLA-ALAAQG 263


RESULT 6
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             8.0%;  Score 245;  DB 9;  Length 249;
  Best Local Similarity   30.1%;  Pred. No. 3.8e-13;
  Matches   81;  Conservative   39;  Mismatches   89;  Indels   60;  Gaps   12;

Qy        341 LLLDGGNSRLKW-------AWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGC-- 391
              | || ||| :||       | :| |       |   ::: |          :||: ||
Db          3 LELDCGNSFIKWRVIHVADAVIEGGGIVDSDQALVAEVAAL---------ASVRLTGCRI 53
```

```
Qy        392 ----------AVCGEFKKA-QVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNA 440
                        |:|    :|  || ::|  :       :  |:|| |   :  | ||| |
Db         54 VSVRSEEETDALCALIAQAFAVQAKVAHPVR------EMAGVRNGYDDYQRLGMDRWLAA 107

Qy        441 LGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGK 500
              ||:   :: ||:|:  |||   | :: || :||| | ||  ||:  |   |  :
Db        108 LGAFHLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRI------ 161

Qy        501 RYPFPTTTGNAVAS-----GMMDAV-CGSVMMMHG----RLKE-KTGAGKPVDVIITGGG 549
              ||    :   |::|       ::||  | |:|: |    :|:: :   |:   | :|||
Db        162 RYD-DASAERALSSLSPGRSTVEAVERGCVLMLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        550 AAKVAEALPPAFLAENTVRVADNLVIYGL 578
              |  |  ||| |       ||  :||  ||
Db        221 APLVRAALPQA-------RVVPDLVFVGL 242


RESULT 7
US-09-813-453A-20
; Sequence 20, Application US/09813453A
; Patent No. US20020168681A1
  GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

  Query Match             7.6%;  Score 234.5;  DB 9;  Length 248;
  Best Local Similarity   27.6%;  Pred. No. 3.1e-12;
  Matches   71;  Conservative   40;  Mismatches  109;  Indels   37;  Gaps    7;

Qy        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400
              | || ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Qy        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453
              :  : :||  :  | ||        |  |:|| |   :  | |||   : :   :: ||:|
Db         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Qy        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Qy        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561
                | | |  :    | ::|: | ::|:        |   :: :||| |  | : |  |
Db        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

Qy        562 LAENTVRVADNLVIYGL 578
                    |:  :||  ||
Db        231 ------RIMPDLVFVGL 241


RESULT 8
US-09-813-453A-65
; Sequence 65, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
;   LENGTH: 242
;   TYPE: PRT
;   ORGANISM: Xylessa fastidiosa
US-09-813-453A-65

  Query Match             6.5%;  Score 200;  DB 9;  Length 242;
  Best Local Similarity   30.6%;  Pred. No. 2.8e-09;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   24;  Gaps    8;

Qy        342 LLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKAQ 401
              | | |||| | | :  |     |    ||   :          |   | | :  :|
Db          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63

Qy        402 VQEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 458
              | | |      :::   : |   |:|  | |||  | ||:   |||  :     :|| ||
Db         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121

Qy        459 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMD 518
              |:|:| |  :| :||| |     ||:::|  |   |    |    |  | :|: ||
Db        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181

Qy        519 AVCG----SVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVR---VAD 571
              |       |:   | || :       | ::: |||   ||  ||  :| :   |   | |
Db        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHRPTLVLD 230

Qy        572 NLVIY 576
                | |:
Db        231 GLAIW 235


RESULT 9
US-09-859-270-2
; Sequence 2, Application US/09859270
; Patent No. US20020048789A1
; GENERAL INFORMATION:
;  APPLICANT: Huang, Jianzhong
;  APPLICANT:  McDevitt, Damien
;  TITLE OF INVENTION: birA
;  FILE REFERENCE: GM10172
;  CURRENT APPLICATION NUMBER: US/09/859,270
;  CURRENT FILING DATE:  2001-05-17
;  PRIOR APPLICATION NUMBER: 09/175,172
;  PRIOR FILING DATE: 1998-10-20
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 311
;   TYPE: PRT
;   ORGANISM: Streptococcus pneumoniae
US-09-859-270-2

  Query Match             6.3%;  Score 194;  DB 10;  Length 311;
  Best Local Similarity   28.7%;  Pred. No. 1.3e-08;
  Matches   77;  Conservative   46;  Mismatches   99;  Indels   46;  Gaps   12;

Qy         76 LGERSGFQTALKHECASSNDEILELARIAPDKAHK--TICVTHLQSKGRGRQGRKWSHRL 133
              | |    : : | |  |:  :    |: | |  |:  |: :   |: ||||  | :
Db         69 LEENLPIKVSFKPETKSTQLD----AKEAIDLGHEANTLYLASYQTAGRGRFQRSFYSPQ 124

Qy        134 GECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGL-DVQIKWPNDLVVGRDKLGGI 192
                |   | :     : |  :| | : : | |  :|:  | | || ||| ||: :   |:|||
```

```
Db        125 GGIYM-TLHLKPNLPYDKLPSYTLLVAGAVYKAIKNLTLIDVDIKWVNDIYLNNHKIGGI 183

Qy        193 LIE---TVRTGGKTVAVVGIGINFVL---PKEVENAASVQSLFQTASRRGNADAAVLLET 246
Db        184 LTEAMTSVETGLVTDIIIGVGINFTIKDFPQELKEKAA--SLFK-------ATAPITRNE 234

Qy        247 LLVEL---------DAVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKG 297
Db        235 LIIEIWRAFFETPAEELLYLYKKQSFI------------LGKEVTFTLE-QKDYKGLAKD 281

Qy        298 VDGQGVLHLETAEGKQT-VVSGEISLRS 324
Db        282 ISENGKLLVQCDNGKEIWLNSGEISLNS 309


RESULT 10
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             6.2%;  Score 191;  DB 9;  Length 249;
  Best Local Similarity   27.8%;  Pred. No. 1.7e-08;
  Matches   72;  Conservative   31;  Mismatches  116;  Indels   40;  Gaps    8;

Qy        341 LLLDGGNSRLKWAWVENGTFATV-GSAPYRDLSPL-------GAEWAE-------KADGN 385
Db          3 LELDCGNSFIKWRIITKSCSTLVSGGVVDSDTALLECLGNLSGAAFSDCRLVSVRSAEET 62

Qy        386 VRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRR 445
Db         63 AKLV-CALADTFSISPVCAAPAPEL---------AGVINGYDDFARLGLDRWLAFVGAYH 112

Qy        446 FSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN------RHAG 499
Db        113 LVKGACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEAEKAL 172

Qy        500 KRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEALPP 559
Db        173 VRLVPGRATAEAVERGCSLMLRGFAMTQIEIAREYW--GDDFAIFVTGGDAVLVADVLPG 230

Qy        560 AFLAENTVRVADNLVIYGL 578
Db        231 A-------RIVPDLVFVGL 242


RESULT 11
US-09-738-626-4286
; Sequence 4286, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4286
;   LENGTH: 288
;   TYPE: PRT
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-4286

  Query Match             6.0%;  Score 184.5;  DB 9;  Length 288;
  Best Local Similarity   30.0%;  Pred. No. 7.8e-08;
  Matches   71;  Conservative   36;  Mismatches  113;  Indels   17;  Gaps    8;

Qy         92 SSNDEILELARI-APDKAHKTICVTHLQSKGRGRQGRKWSHRLGECLMFSFGWVFDRPQY 150
Db         57 STNADLLALAGSGAPNWTVKTV---EFQDHARGRLGRPWSAPEGSQTIVSVLVQLSIDQV 113

Qy        151 E-LGSLSPVAAVACRRALSRLGLD-VQIKWPNDLVVGRDKLGGILIETVRTGGKTVAVVG 208
Db        114 DRIGTIPLAAGLAVMDALNDLGVEGAGLKWPNDVQIHGKKLCGILVEATGFDSTPTVVIG 173

Qy        209 IGINFVLPKE---VENAASVQSLFQTASRRGNADAAVLLETLLVELDAVLLQYARDGFAP 265
Db        174 WGTNISLTKEELPVPHATSL------ALEGVEVDRTTFLINMLTHLHTRLDQW-QGPSVD 226

Qy        266 FVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK-QTVVSGEIS 321
Db        227 WLDDYRAVCSSIGQDVRVLLPGDKELLGEAIGVATGGEIRVRDASGTVHTLNAGEIT 283


RESULT 12
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

  Query Match             5.8%;  Score 177;  DB 9;  Length 265;
  Best Local Similarity   30.5%;  Pred. No. 3.1e-07;
  Matches   60;  Conservative   28;  Mismatches   89;  Indels   20;  Gaps    6;
```

```
Qy        380 EKADGNVRIVGCAVCGEFKKA--QVQEQLARKIEWLPSSAQALGIRNHY----RHPEEHG 433
              |  ||   | | |:|        :::|   |    :|:     |::         ||:| |
Db         56 ELGDG---IDGIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVPILTDHPKEVG 112

Qy        434 SDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTAN 493
              :|| ||: :        :||  ||| | ||::  | |:|| | ||  :  |:| |: |
Db        113 ADRIINAVAAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEISVEALGVKGAQ 172

Qy        494 LNR--HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRL-KEKTGAGKPVDVIITGGGA 550
              |:   |  |      |  |: ||::    | |  :  |: :|       | || |||
Db        173 LRKIEVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDDVTVIATGG-- 230

Qy        551 AKVAEALPPAFLAENTV 567
                  |  |  |::|
Db        231 ------LAPMVLGESSV 241


RESULT 13
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
  GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT:  Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION:  ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE:  2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-813-453A-51

  Query Match             5.6%;  Score 172;  DB 9;  Length 260;
  Best Local Similarity   32.3%;  Pred. No. 8.1e-07;
  Matches   53;  Conservative   22;  Mismatches   81;  Indels    8;  Gaps    3;

Qy        420 LGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG 479
              |||       | | |:||  ||:|:        ||:  ||| | | :  || : || | ||
Db         94 LGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPG 153

Qy        480 FHLMKESLAVRTANLNRHAGKRYPFPTTTGNAVASGMMDAV-CGSVMMMHGRL-KEKTGA 537
              :|  ::|    | | | | :|      |   | |   |  | : :: | : : |
Db        154 INLSMQALHEAAAKLPRIAIQRPAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAER 213

Qy        538 GKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNM 581
              |:|: || ||| |:    |       ::      :| | ||| :
Db        214 GEPMTVIATGGVASLFEGA------TDSIDHFDSDLTIRGLLEI 251


RESULT 14
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT:  Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION:  ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE:  2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 262
;   TYPE: PRT
;   ORGANISM: Bacillus anthracis
US-09-813-453A-45

  Query Match             5.5%;  Score 169.5;  DB 9;  Length 262;
  Best Local Similarity   29.1%;  Pred. No. 1.3e-06;
  Matches   52;  Conservative   27;  Mismatches   77;  Indels   23;  Gaps    4;

Qy        421 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 480
              |:   | :| | |:||  ||:      :  ::|  ||| |   : :: ||:|| | ||
Db         95 GLNIKYENPREVGADRIVNAVAGIHLYGSPLIIVDFGTATTYCYINEEKHYMGGVITPGI 154

Qy        481 HLMKESLAVRTANLNR--------HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLK 532
               :  |:|  | | | |          ||       | :|: ||::    | |  :  |:|
Db        155 MISAEALYSRAAKLPRIEITKPSSVVGK------NTVSAMQSGILYGYVGQVEGIVKRMK 208

Qy        533 EKTGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591
              |:  | :   || ||| |  ::|         | : | |  :    | |:       :|
Db        209 EE--AKQEPKVIATGGLAKLISE-------ESNVIDVVDPPLTLKGLYMLYERNANLQH 258


RESULT 15
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT:  Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION:  ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE:  2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

  Query Match             5.5%;  Score 168;  DB 9;  Length 250;
  Best Local Similarity   24.7%;  Pred. No. 1.7e-06;
  Matches   60;  Conservative   42;  Mismatches   97;  Indels   44;  Gaps    8;

Qy        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKAD--GNVRIVGCAVCGEFK 398
              |:|| ||: :          :|   | | : | |||    |   :    |  |   |:
Db         20 LVLDVGNTNI----------VLGI--YND-TKLTAEWRLSTDVLRSADEYGIQVMNLFQ 65

Qy        399 KAQVQEQLARKI---EWLPSSAQAL--------------------GIRNHYRHPEEHGS 434
              : ::   |   :     :|:   :|                    ||   | :|:| |:
Db         66 QDKLDPTLVEGVIISSVVPNIMYSLEHMIRKYFKINPLVVGPGIKTGINIKYDNPKEVGA 125

Qy        435 DRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANL 494
              ||  ||: :    : : :::  ||| |  |: ::| |||| | ||  :  |:|  : | |
Db        126 DRIVNAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEALFEKAAKL 185

Qy        495 NR--HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGK--PVDVIITGGGA 550
               |       |     | ::: ||::      |  :  :|||    |:     :::  ||
Db        186 PRVELIKPAYAICKNTISSIQSGIVYRYLRQVKYLFEKLKENLPDGRRTRTSLVLATGGL 245

Qy        551 AKV 553
              ||:
Db        246 AKL 248
```

Search completed: June 24, 2003, 22:24:09
Job time : 27.9909 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:26 ; Search time 47.2273 Seconds
(without alignments)
1297.879 Million cell updates/sec

Title: US-09-813-453A-39
Perfect score: 2384
Sequence: 1 MGECLMFSFGWAFDRPQYEL.....LVIHGLLNLIAAEGGESEHA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2384	100.0	460	23	AAU91167	Pantothenate kinas
2	2381	99.9	592	20	AAY38618	Neisseria gonorrho
3	2381	99.9	592	21	AAY74911	Neisseria gonorrhe
4	2355	98.8	455	20	AAY38617	Neisseria gonorrho
5	2355	98.8	455	21	AAY74908	Neisseria gonorrhe
6	2223	93.2	592	20	AAY38616	Neisseria meningit
7	2223	93.2	592	21	AAY74913	Neisseria meningit
8	2223	93.2	592	23	AAU91166	Pantothenate kinas
9	2197	92.2	455	21	AAY74910	Neisseria meningit
10	2197	92.2	592	20	AAY38615	Neisseria meningit
11	2197	92.2	592	21	AAY74912	Neisseria meningit
12	2197	92.2	592	23	AAU91169	Pantothenate kinas
13	1783.5	74.8	389	21	AAY74909	Neisseria meningit
14	927.5	38.9	189	20	AAY38614	Neisseria meningit
15	258.5	10.8	267	23	AAU91162	Bordella pertussis
16	232.5	9.8	248	23	AAU91164	Pantothenate kinas
17	232	9.7	249	23	AAU91178	Pantothenate kinas
18	210	8.8	322	20	AAY31601	Escherichia coli b
19	204	8.6	302	22	AAB88546	Haemophilus influe
20	204	8.6	302	23	AAU91477	Haemophilus influe
21	199.5	8.4	242	23	AAU91180	Pantothenate kinas
22	187	7.8	325	23	ABB48003	Listeria monocytog
23	187	7.8	363	22	ABG28654	Novel human diagno
24	185	7.8	260	23	AAU91173	Pantothenate kinas
25	179	7.5	265	23	AAU91151	Streptomyces coeli
26	177	7.4	249	23	AAU91182	Pantothenate kinas
27	169	7.1	250	23	AAU91150	Clostridium acetob
28	168	7.0	262	23	AAU91170	Pantothenate kinas
29	166	7.0	244	23	AAU91168	Pantothenate kinas
30	161.5	6.8	257	23	AAU91174	Pantothenate kinas
31	161	6.8	282	22	AAU64093	Propionibacterium
32	153	6.4	219	23	AAU91176	Pantothenate kinas
33	153	6.4	269	22	AAB76593	Corynebacterium gl
34	153	6.4	288	22	AAG90532	C glutamicum prote
35	152.5	6.4	246	23	AAU91156	Thermotoga maritim
36	151.5	6.4	237	22	AAB96079	Putative P. abyssi
37	151	6.3	258	22	AAU01243	B. subtilis novel
38	151	6.3	258	23	AAU91149	Bacillus subtilis
39	150.5	6.3	291	22	AAU04901	Micromonospora eve
40	148.5	6.2	250	23	ABB55254	Lactococcus lactis
41	148	6.2	294	22	ABB09064	Thermus caldophilu
42	147.5	6.2	311	21	AAY70846	Streptococcus pneu
43	146.5	6.1	272	22	AAG81225	Mycobacterium tube
44	146.5	6.1	272	23	AAU91152	Mycobacterium tube
45	146.5	6.1	273	23	AAU91157	Treponema pallidum

ALIGNMENTS

RESULT 1
AAU91167
ID AAU91167 standard; Protein; 460 AA.
XX
AC AAU91167;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pantothenate kinase (CoaX) #5.
XX
KW Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS Neisseria gonorrhoeae.
XX
PN WO200216601-A2.
XX
PD 28-FEB-2002.
XX
PF 24-AUG-2001; 2001WO-US26531.
XX
PR 24-AUG-2000; 2000US-227860P.
PR 20-MAR-2001; 2001US-0813453.
XX
PA (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI Yocum RR, Patterson TA;
XX
DR WPI; 2002-269358/31.
DR N-PSDB; ABK54188.
XX
PT Identifying potential antibiotic or antimicrobial agent, comprises
PT contacting composition comprising pantothenate kinase (CoaX) protein

PT with test compound and identifying inhibitor of the CoaX protein -
XX
PS Claim 8; Page 92-94; 128pp; English.
XX
CC The invention describes assays for identifying a (potential) antibiotic
CC comprising contacting an assay composition comprising a pantothenate
CC kinase (CoaX) protein with a test compound, and determining the ability
CC of the test compound to inhibit the activity of the CoaX protein, an
CC essential enzyme for the production of coenzyme A. CoaX protein is a
CC valuable target for identifying bactericidal compounds. CoaX modulating
CC agents can be used in an infectious animal model to determine the
CC efficacy, toxicity, or side effects of treatment with such an agent. This
CC is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC described in the invention.
XX
SQ Sequence 460 AA;

Query Match 100.0%; Score 2384; DB 23; Length 460;
Best Local Similarity 100.0%; Pred. No. 8.5e-209;
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
              |||||||||||||||||||||||||||||||||||||||
Db        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
```

RESULT 2
AAY38618
ID AAY38618 standard; Protein; 592 AA.
XX
AC AAY38618;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigenic protein encoded by ORF61.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI; 1999-327407/27.
DR N-PSDB; AAZ12075.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 175; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 592 AA;

Query Match 99.9%; Score 2381; DB 20; Length 592;
Best Local Similarity 99.8%; Pred. No. 2.3e-208;
Matches 459; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        253 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
              |||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 592
```

RESULT 3
AAY74911
ID AAY74911 standard; Protein; 592 AA.

```
XX
AC   AAY74911;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1296.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53673.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 713; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             99.9%;  Score 2381;  DB 21;  Length 592;
  Best Local Similarity   99.8%;  Pred. No. 2.3e-208;
  Matches  459;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        133 LGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        253 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        493 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
              ||||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 592


RESULT 4
AAY38617
ID   AAY38617 standard; Protein; 455 AA.
XX
AC   AAY38617;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria gonorrhoeae antigen encoded by partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria gonorrhoeae.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 175; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   455 AA;
```

```
  Query Match             98.8%;  Score 2355;  DB 20;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-206;
  Matches  455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 65
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 60

Qy         66 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 120

Qy        126 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 180

Qy        186 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

          246 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        306 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 360

Qy        366 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 420

Qy        426 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
              ||||||||||||||||||||||||||||||||||||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 455


RESULT 5
AAY74908
ID   AAY74908 standard; Protein; 455 AA.
XX
AC   AAY74908;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria gonorrheae ORF 311 protein sequence SEQ ID NO:1290.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
     antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
     antibacterial; gene therapy.
XX
OS   Neisseria gonorrheae.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53670.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 709; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             98.8%;  Score 2355;  DB 21;  Length 455;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-206;
  Matches  455;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          6 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 65
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 60

Qy         66 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 120

Qy        126 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 180

Qy        186 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 245
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        246 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 305
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        306 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 360

Qy        366 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 425
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 420

Qy        426 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
              ||||||||||||||||||||||||||||||||||||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 455


RESULT 6
AAY38616
ID   AAY38616 standard; Protein; 592 AA.
XX
AC   AAY38616;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis strain A antigen encoded by ORF61.
```

```
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12074.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 173; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             93.2%;  Score 2223;  DB 20;  Length 592;
  Best Local Similarity   93.7%;  Pred. No. 6.2e-194;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:|||||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| ||||||||:||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||| |||||||||||||| |||||:||||||||||||||||||||||||||||||
Db        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 7
AAY74913
ID   AAY74913 standard; Protein; 592 AA.
XX
AC   AAY74913;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1300.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR-) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI   Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI   Tettelin H, Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53675.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 715; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             93.2%;  Score 2223;  DB 21;  Length 592;
  Best Local Similarity   93.7%;  Pred. No. 6.2e-194;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;
```

```
Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||| |||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 8
AAU91166
ID   AAU91166 standard; Protein; 592 AA.
XX
AC   AAU91166;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #4.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria meningitidis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein -
XX
PS   Claim 6; Page 85-86; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   592 AA;

  Query Match             93.2%;  Score 2223;  DB 23;  Length 592;
  Best Local Similarity   93.7%;  Pred. No. 6.2e-194;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||| |||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||:|||||||||| ||||||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 9
AAY74910
ID   AAY74910 standard; Protein; 455 AA.
XX
AC   AAY74910;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1294.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
```

```
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53672.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics -
XX
PS   Claim 2; Page 711; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   455 AA;

  Query Match             92.2%;  Score 2197;  DB 21;  Length 455;
  Best Local Similarity   93.8%;  Pred. No. 9.9e-192;
  Matches  426;  Conservative   11;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          6 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 65
              |||||| |||||||||||||||:|||||  |||:||||||||||||||||||||||||
Db          1 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGILIETV 60

Qy         66 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 125
              | |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |
Db         61 RTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELDAVLLQ 120

Qy        126 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 185
              || :|||||: ||: |||||||||||||||||| |||||||||:||||||||:|||||
Db        121 YARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGKQTVVS 180

Qy        186 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 245
              |||||| |:| |||||| ||||||||:|||||||||||||||||||||||||||||||
Db        181 GEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 240

Qy        246 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 305
              |||| |||||||||||||| |||||:|||||||||||||||||||||||||||||||||
Db        241 WAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRW 300

Qy        306 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 365
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRH 360

Qy        366 AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEAL 425
              |||||||||||||||||||||||||||:|||||||| ||||||||||||||||||||||
Db        361 AGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEAL 420

Qy        426 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||||||||||||
Db        421 PPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 454

RESULT 10
AAY38615
ID   AAY38615 standard; Protein; 592 AA.
XX
AC   AAY38615;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS   Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12073.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 172; 524pp; English.
XX
CC   Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC   and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   592 AA;

  Query Match             92.2%;  Score 2197;  DB 20;  Length 592;
  Best Local Similarity   92.6%;  Pred. No. 1.5e-191;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||:|||||  |||: |||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||||||||||||||||||| ||||||:|||||||||||||||||||||||||||
```

```
Db        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||:|||||||||| ||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||:||||:|||||  | ||
Db        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 11
AAY74912
ID   AAY74912 standard; Protein; 592 AA.
XX
AC   AAY74912;

     21-MAR-2000  (first entry)

DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1298.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
     Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
     Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53674.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 714; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   592 AA;

  Query Match             92.2%;  Score 2197;  DB 21;  Length 592;
  Best Local Similarity   92.6%;  Pred. No. 1.5e-191;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| ||||||||||||||:|||||  |||: ||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| ||||||||:||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||||||||||||||||||||||||| |||||:|||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||:||||:|||||  | ||
Db        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 12
AAU91169
ID   AAU91169 standard; Protein; 592 AA.
XX
AC   AAU91169;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Pantothenate kinase (CoaX) #7.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Neisseria meningitidis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
DR   N-PSDB; ABK54190.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
```

```
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 96-98; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   592 AA;

 Query Match             92.2%;  Score 2197;  DB 23;  Length 592;
 Best Local Similarity   92.6%;  Pred. No. 1.5e-191;
 Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:||||||  |||: |||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||||||||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||:|||||||| ||||||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||:||||:||||| | ||
Db        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 13
AAY74909
ID   AAY74909 standard; Protein; 389 AA.
XX
AC   AAY74909;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 311 protein sequence SEQ ID NO:1292.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ53671.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 710; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   389 AA;

 Query Match             74.8%;  Score 1783.5;  DB 21;  Length 389;
 Best Local Similarity   78.5%;  Pred. No. 4.4e-154;
 Matches  357;  Conservative    7;  Mismatches   24;  Indels   67;  Gaps    2;

Qy          6 MFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETV 65
              |||||| ||||||||||||||||:|||||  |||: |||||||||||||||||||||
Db          1 MFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGILIETV 60

Qy         66 RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQ 125
              | |||||||||||||||| ||||||||||||||||||||||||||||
Db         61 RTGGKTVAVVGIGINFVLPXEVENAASVQSLFQTASRRGNADAAVLLXX---------- 109

Qy        126 YAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVS 185

Db        110 ----------------------------------------------------XXXXX 114

Qy        186 GEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 245
                  ||||| | | ||| || ||||||||:||||||||||||||||||||||||||||||
Db        115 XEISLRSDXRPVSVXKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAE 174

Qy        246 WAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQAL-GIRNHYRHPEEHGSDR 304
              |||||||||||||||||| ||||:|||||||||||||||| ||||||||||||||||
Db        175 WAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALFGIRNHYRHPEEHGSDR 234

Qy        305 WFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 WFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNR 294

Qy        365 PAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEA 424
```

```
                 |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||||
Db        295 HAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAKVAEA 354

Qy        425 LPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
                 ||||||||||||||||||:||||:||||| | ||
Db        355 LPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 389


RESULT 14
AAY38614
ID   AAY38614 standard; Protein; 189 AA.
XX
AC   AAY38614;
XX
DT   08-OCT-1999  (first entry)
XX
DE   Neisseria meningitidis antigen encoded by a partial ORF61.
XX
KW   Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW   treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
     Neisseria meningitidis.
XX
PN   WO9924578-A2.
XX
PD   20-MAY-1999.
XX
PF   09-OCT-1998;   98WO-IB01665.
XX
PR   01-SEP-1998;   98GB-0019016.
PR   06-NOV-1997;   97GB-0023516.
PR   14-NOV-1997;   97GB-0024190.
PR   18-NOV-1997;   97GB-0024386.
PR   27-NOV-1997;   97GB-0025158.
PR   10-DEC-1997;   97GB-0026147.
PR   14-JAN-1998;   98GB-0000759.
XX
PA   (CHIR-) CHIRON SPA.
XX
PI   Grandi G,  Masignani V,  Pizza M,  Rappuoli R,  Scarlato V;
XX
DR   WPI; 1999-327407/27.
DR   N-PSDB; AAZ12072.
XX
PT   Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT   diagnosis, treatment and prevention of infection
XX
PS   Claim 4; Page 171; 524pp; English.
XX
     Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
     and N. gonorrhoeae antigenic proteins. They are encoded by open
CC   reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC   their fragments, their nucleic acids and antibodies are used for
CC   diagnosis, prevention (as vaccines) or treatment of Neisseria
CC   infections, such as meningitis, septicaemia and gonorrhea. Both
CC   organisms are closely related. Fragments of the nucleic acids
CC   are useful as hybridisation probes and antisense reagents.
XX
SQ   Sequence   189 AA;

  Query Match             38.9%;  Score 927.5;  DB 20;  Length 189;
  Best Local Similarity   94.2%;  Pred. No. 2e-76;
  Matches  178;  Conservative    2;  Mismatches    8;  Indels    1;  Gaps    1;

Qy        187 EISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEW 246
                 ||||| | | ||| || ||||||||:||||||||||||||||||||||||||||||||
Db          1 EISLRSDXRPVSVXKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEW 60

Qy        247 AEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWF 306
                 |||||||||||||||| |||||:|||||||||||||| ||||||||||||||||||
Db         61 AEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQAXGIRNHYRHPEEHGSDRWF 120

Qy        307 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLG-GTIMPGFHLMKESLAVRTANLNRP 365
                 |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        121 NALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGXGTIMPGFHLMKESLAVRTANLNRH 180

Qy        366 AGKRYPFPT 374
                 |||||||||
Db        181 AGKRYPFPT 189


RESULT 15
AAU91162
ID   AAU91162 standard; Protein; 267 AA.
XX
AC   AAU91162;
XX
DT   05-JUN-2002  (first entry)
XX
DE   Bordella pertussis pantothenate kinase CoaX.
XX
KW   Pantothenate kinase; CoaX; antibiotic; antimicrobial;
KW   pantothenate kinase modulator; coenzyme A; bactericidal compound.
XX
OS   Bordetella pertussis.
XX
PN   WO200216601-A2.
XX
PD   28-FEB-2002.
XX
PF   24-AUG-2001; 2001WO-US26531.
XX
PR   24-AUG-2000; 2000US-227860P.
PR   20-MAR-2001; 2001US-0813453.
XX
PA   (OMNI-) OMNIGENE BIOPRODUCTS INC.
XX
PI   Yocum RR,  Patterson TA;
XX
DR   WPI; 2002-269358/31.
XX
PT   Identifying potential antibiotic or antimicrobial agent, comprises
PT   contacting composition comprising pantothenate kinase (CoaX) protein
PT   with test compound and identifying inhibitor of the CoaX protein  -
XX
PS   Claim 6; Page 79-80; 128pp; English.
XX
CC   The invention describes assays for identifying a (potential) antibiotic
CC   comprising contacting an assay composition comprising a pantothenate
CC   kinase (CoaX) protein with a test compound, and determining the ability
CC   of the test compound to inhibit the activity of the CoaX protein, an
CC   essential enzyme for the production of coenzyme A. CoaX protein is a
CC   valuable target for identifying bactericidal compounds. CoaX modulating
CC   agents can be used in an infectious animal model to determine the
CC   efficacy, toxicity, or side effects of treatment with such an agent. This
CC   is the amino acid sequence of a pantothenate kinase (CoaX) protein
CC   described in the invention.
XX
SQ   Sequence   267 AA;

  Query Match             10.8%;  Score 258.5;  DB 23;  Length 267;
  Best Local Similarity   31.8%;  Pred. No. 4.1e-15;
  Matches   84;  Conservative   43;  Mismatches  116;  Indels   21;  Gaps    8;

Qy        209 LLLEGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 263
                 :|:: ||||||  | : :   |    ||      ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        264 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 317
                 ::  :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        318 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTG 377
                 :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182
```

```
Qy        378 NAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGG-------AAKVAEALPPAFL 430
              |:|||:  |  |:|:     |  :   |:  :: : |||       | ::       ||
Db        183 QAIASGIAAAQAGAIVRQW--LAGRQRYGQAPEIYVAGGGWPEVRQEAERLLAVTGAAFG 240

Qy        431 AENTVRVADNLVIHGLLNLIAAEG 454
              |       |: |: ||  | ||:|
Db        241 ATPQPTYLDSPVLDGLAAL-AAQG 263
```

Search completed: June 24, 2003, 21:46:28
Job time : 48.2273 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:33:11 ; Search time 8.00068 Seconds
(without alignments)
2384.688 Million cell updates/sec

Title: US-09-813-453A-39
Perfect score: 2384
Sequence: 1 MGECLMFSFGWAFDRPQYEL.....LVIHGLLNLIAAEGGESEHA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	258.5	10.8	267	1	BAF_BORPE	Q45338 bordetella
2	210	8.8	321	1	BIRA_ECOLI	P06709 escherichia
3	204	8.6	302	1	BIRA_HAEIN	P46363 haemophilus
4	188	7.9	320	1	BIRA_SALTY	P37416 salmonella
5	167	7.0	240	1	BIRA_PARDE	P29906 paracoccus
6	161	6.8	325	1	BIRA_BACSU	P42975 bacillus su
7	134	5.6	233	1	YACB_BACSU	P37564 bacillus su
8	132.5	5.6	1655	1	OMPB_RICCN	Q9kka3 r outer mem
9	128	5.4	1656	1	OMPB_RICJA	O06653 r outer mem
10	118	4.9	1654	1	OMPB_RICRI	Q53047 r outer mem
11	111.5	4.7	847	1	RELA_STRCO	P52560 streptomyce
12	111	4.7	548	1	FUMB_ECOLI	P14407 escherichia
13	105	4.4	417	1	SAHH_SULSO	P50252 sulfolobus
14	104	4.4	841	1	RELA_STRAT	O85709 streptomyce
15	103.5	4.3	541	1	CH60_PANAN	O66218 pantoea ana
16	103	4.3	540	1	CH60_ERWCA	O66220 erwinia car
17	102.5	4.3	495	1	MURE_XYLFA	Q9pf85 xylella fas
18	102.5	4.3	726	1	BPL1_HUMAN	P50747 h biotin--p
19	101.5	4.3	1756	1	TRI2_ECOLI	P22706 escherichia
20	101	4.2	547	1	FUMA_ECOLI	P00923 escherichia
21	100.5	4.2	775	1	PLSB_MYCLE	Q9x7b0 mycobacteri
22	100	4.2	507	1	ATPA_CHLRE	P26526 chlamydomon
23	100	4.2	573	1	CH60_DROME	O02649 drosophila
24	99.5	4.2	1110	1	CYGD_BOVIN	P55203 bos taurus
25	99.5	4.2	1645	1	OMPB_RICTY	P96989 r outer mem
26	99	4.2	916	1	HMDH_DROME	P14773 drosophila
27	99	4.2	1132	1	VHSJ_LAMBD	P03749 bacteriopha
28	98.5	4.1	1325	1	MRP4_HUMAN	O15439 homo sapien
29	98	4.1	417	1	PGKB_LEIMA	Q27683 leishmania
30	98	4.1	437	1	SAHH_PYRAE	Q8ztq7 pyrobaculum
31	98	4.1	447	1	AROA_SYNY3	Q59975 synechocyst
32	98	4.1	479	1	PGKC_LEIMA	P50312 leishmania
33	98	4.1	596	1	DHSA_RICCN	Q92j97 rickettsia
34	98	4.1	1276	1	PMP6_CHLPN	Q9z899 chlamydia p
35	98	4.1	1302	1	ACSC_ACEXY	P37718 acetobacter
36	97.5	4.1	356	1	FRVX_ECOLI	P32153 escherichia
37	97.5	4.1	540	1	CH60_ERWAP	O66222 erwinia aph
38	97	4.1	479	1	PGKC_LEIME	Q27685 leishmania
39	97	4.1	506	1	ATPA_EUGGR	P30392 euglena gra
40	97	4.1	529	1	ATPB_MOUSE	P56480 mus musculu
41	95.5	4.0	806	1	HYF2_ALCEU	O07451 alcaligenes
42	95.5	4.0	989	1	PTP3_DICDI	P54637 dictyosteli
43	95	4.0	505	1	ATPB_KLULA	P49376 kluyveromyc
44	95	4.0	579	1	FUMA_SALTY	P40720 salmonella
45	95	4.0	691	1	YHJG_ECOLI	P37645 escherichia

ALIGNMENTS

RESULT 1
BAF_BORPE
ID BAF_BORPE STANDARD; PRT; 267 AA.
AC Q45338; Q45373;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Bvg accessory factor.
GN BAF.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC Bordetella.
OX NCBI_TaxID=520;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP504;
RX MEDLINE=95325323; PubMed=7601846;
RA Deshazer D., Wood G.E., Friedman R.L.;
RT "Identification of a Bordetella pertussis regulatory factor required
RT for transcription of the pertussis toxin operon in Escherichia
RT coli.";
RL J. Bacteriol. 177:3801-3807(1995).
RN [2]
RP SEQUENCE OF 1-38 FROM N.A.
RC STRAIN=BP504;
RA Wood G.E., Friedman R.L.;
RT "Identification of a birA homolog in Bordetella pertussis.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 239-267 FROM N.A.
RC STRAIN=BP536;
RX MEDLINE=96419162; PubMed=8821935;
RA Allen A.G., Maskell D.J.;
RT "The identification, cloning and mutagenesis of a genetic locus
RT required for lipopolysaccharide biosynthesis in Bordetella
RT pertussis.";
RL Mol. Microbiol. 19:37-52(1996).
CC -!- FUNCTION: ACTIVATES TRANSCRIPTION OF THE PERTUSSIS TOXIN OPERON IN
CC A BVGAS-DEPENDENT MANNER. MAY INTERACT WITH THE ALPHA SUBUNIT OF
CC RNA POLYMERASE.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U12020; AAA75361.1; -.
DR EMBL; AF016461; AAC68834.1; -.
DR EMBL; X90711; CAA62242.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Transcription regulation; Activator.
SQ SEQUENCE 267 AA; 27845 MW; 336A615F67B57901 CRC64;

```
  Query Match             10.8%;  Score 258.5;  DB 1;  Length 267;
  Best Local Similarity   31.8%;  Pred. No. 1.5e-11;
  Matches   84;  Conservative   43;  Mismatches  116;  Indels   21;  Gaps    8;

Qy        209 LLLEGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 263
              :|:: ||||||  | : :   |    ||      ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        264 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 317
              ::   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        318 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTG 377
              :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        378 NAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGG-------AAKVAEALPPAFL 430
              |:|||:  |  |:|:     |  :   |:  :: : |||       | ::      ||
Db        183 QAIASGIAAAQAGAIVRQW--LAGRQRYGQAPEIYVAGGGWPEVRQEAERLLAVTGAAFG 240

Qy        431 AENTVRVADNLVIHGLLNLIAAEG 454
              |       |: |: ||  | ||:|
Db        241 ATPQPTYLDSPVLDGLAAL-AAQG 263


RESULT 2
BIRA_ECOLI
ID   BIRA_ECOLI     STANDARD;      PRT;   321 AA.
AC   P06709;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR BIOR OR DHBB OR B3973.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86006282; PubMed=3899863;
RA   Howard P.K., Shaw J., Otsuka A.J.;
RT   "Nucleotide sequence of the birA gene encoding the biotin operon
RT   repressor and biotin holoenzyme synthetase functions of Escherichia
RT   coli.";
RL   Gene 35:321-331(1985).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87055242; PubMed=3536662;
RA   Buoncristiani M.R., Howard P.K., Otsuka A.J.;
RT   "DNA-binding and enzymatic domains of the bifunctional biotin operon
RT   repressor (BirA) of Escherichia coli.";
RL   Gene 44:255-261(1986).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
RN   [4]
RP   SEQUENCE OF 1-76 FROM N.A.
RC   STRAIN=RDD012;
RX   MEDLINE=92165747; PubMed=1311302;
RA   Pucci M.J., Discotto L.F., Dougherty T.J.;
RT   "Cloning and identification of the Escherichia coli murB DNA
RT   sequence, which encodes UDP-N-acetylenolpyruvoylglucosamine
RT   reductase.";
RL   J. Bacteriol. 174:1690-1693(1992).
RN   [5]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=93028443; PubMed=1409631;
RA   Wilson K.P., Shewchuk L.M., Brennan R.G., Otsuka A.J., Matthews B.W.;
RT   "Escherichia coli biotin holoenzyme synthetase/bio repressor crystal
RT   structure delineates the biotin- and DNA-binding domains.";
RL   Proc. Natl. Acad. Sci. U.S.A. 89:9257-9261(1992).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M10123; AAA23520.1; -.
DR   EMBL; M15820; AAA23521.1; -.
DR   EMBL; L14557; AAA24186.1; -.
DR   EMBL; U00006; AAC43075.1; -.
DR   EMBL; AE000471; AAC76951.1; -.
DR   PIR; B24029; BVECBF.
DR   PIR; A24989; A24989.
DR   PDB; 1BIA; 31-OCT-93.
DR   PDB; 1BIB; 31-OCT-93.
DR   EcoGene; EG10123; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   3D-structure; Complete proteome.
FT   DNA_BIND      22     41       H-T-H MOTIF.
FT   VARIANT       61     61       P -> A (IN STRAIN RDD012).
FT   VARIANT       70     70       K -> E (IN STRAIN RDD012).
FT   HELIX          5     14
FT   TURN          15     16
FT   STRAND        20     20
FT   HELIX         22     29
FT   TURN          30     30
FT   HELIX         33     45
FT   TURN          46     47
FT   STRAND        51     53
FT   TURN          54     56
FT   STRAND        57     59
FT   HELIX         69     74
FT   TURN          75     75
FT   STRAND        81     83
FT   STRAND        86     87
FT   HELIX         90    100
FT   TURN         103    104
FT   STRAND       106    109
FT   STRAND       111    112
FT   TURN         127    128
FT   STRAND       130    139
FT   TURN         146    146
FT   HELIX        147    163
```

```
FT   TURN        164    165
FT   TURN        167    168
FT   STRAND      170    172
FT   TURN        173    175
FT   STRAND      176    179
FT   TURN        180    181
FT   STRAND      182    192
FT   STRAND      199    208
FT   TURN        226    226
FT   HELIX       227    229
FT   TURN        230    230
FT   HELIX       235    254
FT   TURN        255    256
FT   HELIX       259    268
FT   STRAND      269    269
FT   TURN        272    273
FT   STRAND      275    280
FT   TURN        281    282
FT   STRAND      283    292
FT   TURN        294    295
FT   STRAND      298    302
FT   TURN        303    304
FT   STRAND      305    309
FT   STRAND      313    313
FT   STRAND      316    316
SQ   SEQUENCE   321 AA;  35312 MW;  B80AEBCEEE1BD2D4 CRC64;

  Query Match             8.8%;  Score 210;  DB 1;  Length 321;
  Best Local Similarity   33.7%;  Pred. No. 5.4e-08;
  Matches   65;  Conservative   29;  Mismatches   93;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| ||| : : || :  |          | |   |   |: |
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAGINLDRNTLAAMLIRE 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| ||:|: :|  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLAPYLSRWEKLDNFINRPVKLI-IGDKEIFGISRGIDKQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 IIKPWMGGEISLR 317
```

```
RESULT 3
BIRA_HAEIN
ID   BIRA_HAEIN     STANDARD;      PRT;   302 AA.
AC   P46363;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR HI0220.1.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Rd / KW20 / ATCC 51907;
RX   MEDLINE=95350630; PubMed=7542800;
RA   Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA   Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA   McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA   Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA   Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA   Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA   Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA   Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA   Venter J.C.;
RT   "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
RN   [2]
RP   IDENTIFICATION.
RA   Koonin E.V., Rudd K.E.;
RL   Submitted (SEP-1995) to the SWISS-PROT data bank.
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32708; AAC21889.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; HI0220.1; -.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
SQ   SEQUENCE   302 AA;  34341 MW;  50EE9D6E9820378B CRC64;

  Query Match             8.6%;  Score 204;  DB 1;  Length 302;
  Best Local Similarity   32.3%;  Pred. No. 1.3e-07;
  Matches   61;  Conservative   27;  Mismatches   89;  Indels   12;  Gaps    6;

Qy          5 LMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIET 64
              :|||| |||| |:  :  || |  ||          |  |:||||::    ||||||:|
Db        122 IMFSFYWAFD-PKKSIEGLSLVIGLAIAEV-----LNVQVKWPNDILFDERKLGGILVEI 175

Qy         65 V-RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVL 123
                 |    |:|||||  | |: | :   |   :       : :   ||  |:  |   |
Db        176 ANHKNGMLNLVIGIGINVSLSKQTEIS---QPYAEVCEIDPDVERQTLLPKLIQHLYTRL 232

Qy        124 EQYAEEGF-APFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQT 182
              : : |    |   ::: |      : :| : : |  |  :|:| || | :      |
Db        233 NIFEQNGIDEEFQQAWQSYNAFSNSEINVLTE-QGVISGIEQGIDERGYLKVLCGNKIQM 291

Qy        183 VVSGEISLR 191
                ||:|||
Db        292 FNGGEVSLR 300


RESULT 4
BIRA_SALTY
ID   BIRA_SALTY     STANDARD;      PRT;   320 AA.
AC   P37416; Q9L9K4;
DT   01-OCT-1994 (Rel. 30, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
```

```
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA OR STM4138 OR STMF1.2.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
RN   [2]
RP   SEQUENCE OF 1-135 FROM N.A.
RC   STRAIN=LT2;
RX   MEDLINE=94324848; PubMed=8048842;
RA   Dombrosky P.M., Schmid M.B., Young K.D.;
RT   "Sequence divergence of the murB and rrfB genes from Escherichia coli
RT   and Salmonella typhimurium.";
RL   Arch. Microbiol. 161:501-507(1994).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SUBUNIT: MONOMER.
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF170176; AAF33493.1; -.
DR   EMBL; AE008893; AAL22971.1; -.
DR   EMBL; L14816; -; NOT_ANNOTATED_CDS.
DR   HSSP; P06709; 1BIA.
DR   StyGene; SG10028; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
FT   DNA_BIND     22     41       H-T-H MOTIF.
FT   CONFLICT    100    101       EL -> DV (IN REF. 2).
SQ   SEQUENCE   320 AA;  35395 MW;  010E8483093297F5 CRC64;

  Query Match             7.9%;  Score 188;  DB 1;  Length 320;
  Best Local Similarity   31.6%;  Pred. No. 2e-06;
  Matches   61;  Conservative   32;  Mismatches   94;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLLDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| |||  : : ||     |           |  :|   |: :
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGITLDRNMLAAKLIYK 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| :|:|: ::  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLI-IGDKEIFGISRGIDTQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 VIKPWMGGEISLR 317


RESULT 5
BIRA_PARDE
ID   BIRA_PARDE     STANDARD;      PRT;   240 AA.
AC   P29906;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15) (Biotin--
DE   protein ligase).
GN   BIRA.
OS   Paracoccus denitrificans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Paracoccus.
OX   NCBI_TaxID=266;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 13543;
RX   MEDLINE=93136200; PubMed=8422400;
RA   Xu X., Matsuno-Yagi A., Yagi T.;
RT   "DNA sequencing of the seven remaining structural genes of the gene
RT   cluster encoding the energy-transducing NADH-quinone oxidoreductase
RT   of Paracoccus denitrificans.";
RL   Biochemistry 32:968-981(1993).
CC   -!- FUNCTION: ACTIVATES BIOTIN TO FORM BIOTINYL-5'-ADENYLATE AND
CC       TRANSFERS THE BIOTIN MOIETY TO BIOTIN-ACCEPTING PROTEINS.
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L02354; AAA25601.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Biotin; Ligase.
SQ   SEQUENCE   240 AA;  24407 MW;  9FD9184ACC1221D8 CRC64;

  Query Match             7.0%;  Score 167;  DB 1;  Length 240;
  Best Local Similarity   31.7%;  Pred. No. 4.4e-05;
  Matches   60;  Conservative   22;  Mismatches   85;  Indels   22;  Gaps    8;

Qy         15 RPQ---YELGSLSPVAALACRRALGCL---GLETQIKWPNDLVVGRDKLGGILIETVRAG 68
              |||       || |||||   |||          ||||||:::   |: |||:|:  :|
Db         55 RPQGGALAAAQLSFVAALALYDALGLACGPAARLAIKWPNDVLLNGGKVAGILLESSGSG 114
```

```
Qy         69 -GKTVAVVGIGINFVLPKEVENAASVQ-SLFQTASRRGNADAAV-------LLETLLAEL 119
              |.   ||||:|        :| :|:       | :|    ||       ||      |
Db        115 PGVQAVAVGIGVNLA---GAPDAGAVEPGATPPVSVQGETGHAVDPEEFLDLLAPAFARW 171

Qy        120 GAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEG 179
              | |: |   ||||  | :       |: : : | |     |  :|:|  | | |    |
Db        172 QAQLDTY---GFAPIRNAWLARAARLGEPI-IARTGTAESHGIFEGIDDSGALILRGPAG 227

Qy        180 EQTVVSGEI 188
              | : : |:
Db        228 RQVIPAAEV 236


RESULT 6
BIRA_BACSU
ID   BIRA_BACSU     STANDARD;      PRT;   325 AA.
AC   P42975;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   BirA bifunctional protein [Includes: Biotin operon repressor;
DE   Biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15)
DE   (Biotin--protein ligase)].
GN   BIRA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg, and W23 / PY79;
RX   MEDLINE=95247697; PubMed=7730294;
RA   Bower S., Perkins J.P., Yocum R.R., Serror P., Sorokin A.V.,
RA   Rahaim P., Howitt C.L., Prasad N., Ehrlich S.D., Pero J.;
RT   "Cloning and characterization of the Bacillus subtilis birA gene
RT   encoding a repressor of the biotin operon.";
RL   J. Bacteriol. 177:2572-2575(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168 / Marburg;
RX   MEDLINE=96349105; PubMed=8760912;
RA   Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA   Serror P.;
RT   "Sequence analysis of the Bacillus subtilis chromosome region between
RT   the serA and kdg loci cloned in a yeast artificial chromosome.";
RL   Microbiology 142:2005-2016(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- FUNCTION: BIRA ACTS BOTH AS A BIOTIN-OPERON REPRESSOR AND AS
CC       THE ENZYME THAT SYNTHESIZES THE COREPRESSOR, ACETYL-COA:CARBON-
CC       DIOXIDE LIGASE. THIS PROTEIN ALSO ACTIVATES BIOTIN TO FORM
CC       BIOTINYL-5'-ADENYLATE AND TRANSFERS THE BIOTIN MOIETY TO
CC       BIOTIN-ACCEPTING PROTEINS (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: ATP + biotin + apo-[acetyl-CoA:carbon-dioxide
CC       ligase (ADP forming)] = AMP + diphosphate + [acetyl-CoA:carbon-
CC       dioxide ligase (ADP forming)].
CC   -!- SIMILARITY: WITH OTHER BACTERIAL BIRA AND WITH EUKARYOTIC BIOTIN
CC       APO-PROTEIN LIGASE.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L38424; AAA92879.1; ALT_SEQ.
DR   EMBL; U20445; AAB60184.1; -.
DR   EMBL; L47709; AAB38447.1; -.
DR   EMBL; Z99115; CAB14160.1; -.
DR   HSSP; P06709; 1BIB.
DR   SubtiList; BG11206; birA.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Biotin; Transcription regulation; DNA-binding; Repressor; Ligase;
KW   Complete proteome.
FT   DNA_BIND      23     42       H-T-H MOTIF (BY SIMILARITY).
FT   VARIANT       18     18       N -> S (IN STRAIN W23).
FT   VARIANT       89     89       H -> Q (IN STRAIN W23).
FT   VARIANT       92     92       L -> I (IN STRAIN W23).
FT   VARIANT      137    137       V -> I (IN STRAIN W23).
FT   VARIANT      169    169       E -> A (IN STRAIN W23).
FT   VARIANT      219    220       PN -> SD (IN STRAIN W23).
FT   VARIANT      237    237       A -> S (IN STRAIN W23).
FT   VARIANT      311    311       N -> Q (IN STRAIN W23).
FT   MUTAGEN       23     23       G->S: DEREGULATION OF BIOTIN SYNTHESIS.
FT   MUTAGEN       38     38       W->R: DEREGULATION OF BIOTIN SYNTHESIS.
FT   MUTAGEN       58     58       G->E: DEREGULATION OF BIOTIN SYNTHESIS.
SQ   SEQUENCE   325 AA;  36182 MW;  69088BB2CAD09221 CRC64;

  Query Match             6.8%;  Score 161;  DB 1;  Length 325;
  Best Local Similarity   29.1%;  Pred. No. 0.00017;
  Matches   59;  Conservative   30;  Mismatches   64;  Indels   50;  Gaps    9;

Qy         14 DRPQYELGSLSPVAALACRRAL-GCLGLETQIKWPNDLVVGRDKLGGILIE------TVR 66
              | |  :   |: :|:|:|  : :    |::| ||||||:::   |  ||| |       ||
Db        146 DIPLQKTPQLTLLAAVAVVQGIEEAAGIQTDIKWPNDILINGKKTVGILTEMQAEEDRVR 205

Qy         67 AGGKTVAVVGIGINF-----VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGA 121
              :      ::|||||        | |::: |:  || | |  :   | | ::: :|
Db        206 S-----VIIGIGINVNQQPNDFPDELKDIAT--SLSQAAGEK--IDRAGVIQHILLCFEK 256

Qy        122 VLEQYAEEGFAPFLNEYE------------TANRD-HGKAVLLLRDGETVCEGTVKGVD 167
              |   || |    :|           | |   :|||:               |:|
Db        257 RYRDYMTHGFTPIKLLWESYALGIGTNMRARTLNGTFYGKAL--------------GID 301

Qy        168 GRGVLHLETAEGEQTVVSGEISL 190
```

```
                ||| ||| || : : | :| |
Db        302 DEGVLLLETNEGIKKIYSADIEL 324


RESULT 7
YACB_BACSU
ID   YACB_BACSU     STANDARD;      PRT;   233 AA.
AC   P37564;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein yacB.
GN   YACB.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=96051385; PubMed=7584024;
RA   Ogasawara N., Nakai S., Yoshikawa H.;
RT   "Systematic sequencing of the 180 kilobase region of the Bacillus
RT   subtilis chromosome containing the replication origin.";
RL   DNA Res. 1:1-14(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the Gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
CC   -!- SIMILARITY: SOME, TO B.PERTUSSIS BVG ACCESSORY FACTOR.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26185; BAA05305.1; -.
DR   EMBL; Z99104; CAB11846.1; -.
DR   SubtiList; BG10133; yacB.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   233 AA;  26217 MW;  AAE96E732C15DF44 CRC64;

  Query Match             5.6%;  Score 134;  DB 1;  Length 233;
  Best Local Similarity   31.7%;  Pred. No. 0.0098;
  Matches   38;  Conservative   20;  Mismatches   48;  Indels   14;  Gaps    3;

Qy        289 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 348
              |:   | :|:| |:||  ||: :     |  :|| ||| |   : ::  |:|| | :||
Db         95 GLNIKYDNPKEVGADRIVNAVAAIHLYGNPLIVVDFGTATTYCYIDENKQYMGGAIAPGI 154

Qy        349 HLMKESLAVRTANL-----NRP---AGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLK 400
               :  |:|  | | |      ||    ||       | :|: ||::    | :  :  |:|
Db        155 TISTEALYSRAAKLPRIEITRPDNIIGK------NTVSAMQSGILFGYVGQVEGIVKRMK 208


RESULT 8
OMPB_RICCN
ID   OMPB_RICCN     STANDARD;      PRT;  1655 AA.
AC   Q9KKA3; Q9KK98; Q9XC45;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB OR RC1085.
OS   Rickettsia conorii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=781;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Malish 7;
RX   MEDLINE=21442074; PubMed=11557893;
RA   Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA   Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA   Raoult D.;
RT   "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL   Science 293:2093-2098(2001).
RN   [2]
RP   SEQUENCE OF 33-1649 FROM N.A.
RC   STRAIN=Indian tick typhus, and Malish 7;
RX   MEDLINE=20393643; PubMed=10939649;
RA   Roux V., Raoult D.;
RT   "Phylogenetic analysis of members of the genus Rickettsia using the
RT   gene coding the outer-membrane protein rOmpB (ompB).";
RL   Int. J. Syst. Evol. Microbiol. 50:1449-1455(2000).
RN   [3]
RP   SEQUENCE OF 353-1655 FROM N.A.
RC   STRAIN=Malish 7;
RA   Stenos J., Walker D.;
RT   "The rickettsial outer membrane protein A and B genes of Rickettsia
RT   australis, the most divergent rickettsia of the spotted fever group.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       SIMILARITY).
CC   -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE008659; AAL03623.1; -.
DR   EMBL; AF123721; AAF34124.1; -.
DR   EMBL; AF123726; AAF34129.1; -.
DR   EMBL; AF149110; AAD39533.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall; Complete proteome.
FT   CHAIN         1   1334       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1335   1655       32 KDA BETA PEPTIDE.
FT   VARIANT      61     61       P -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      75     75       G -> S (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT      78     78       K -> N (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     251    251       V -> A (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     413    413       N -> D (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     959    959       I -> V (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT     988    988       A -> T (IN STRAIN INDIAN TICK TYPHUS).
FT   VARIANT    1139   1139       R -> L (IN STRAIN INDIAN TICK TYPHUS).
FT   CONFLICT    353    354       KD -> GH (IN REF. 3).
FT   CONFLICT    776    776       F -> S (IN REF. 3).
FT   CONFLICT   1159   1159       E -> D (IN REF. 3).
FT   CONFLICT   1177   1177       G -> S (IN REF. 3).
FT   CONFLICT   1492   1492       H -> R (IN REF. 3).
SQ   SEQUENCE   1655 AA;  168342 MW;  E49E19377D5FCE37 CRC64;

  Query Match             5.6%;  Score 132.5;  DB 1;  Length 1655;
  Best Local Similarity   21.3%;  Pred. No. 0.14;
  Matches  100;  Conservative   64;  Mismatches  181;  Indels  125;  Gaps   20;

Qy         49 DLVVGRDKLGGILIETVR---------AGGKTVAVVGIGINFVLPKEVENAASVQSLFQT 99
              |: |  |   | :  |             |||: : | ||  |     :|| :| :
Db        101 DVTVREDTTLGFITNVVNNANHFNLMLNAGKTLTITGQGITNVQAAATKNANNVVAQVNN 160

Qy        100 ASRRGNADAAVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVC 159
                :  | |        |  :| :     : | |     :  ||    || |:| |   :
Db        161 GAAIDNND--------LQGVGRI-----DCGAAASTLVFNLANPTTQKAPLILGDNAVIV 207

Qy        160 EG-------------------TVKGV---DGRGVLHLETAEGEQTVVSGEISLRPDNRS 196
               |                   ||  :   ||:|::    |:   |     ::|: :  :
Db        208 NGANGTLNVTNGFIKVSSKSFATVNVINIGDGQGIMFNTDADNVNT-----LNLQANGAT 262

Qy        197 VSVPKRPDSERFLLLE--------------GGNSRLKWAWVENGTFATVGSAPYRDLSPL 242
              ::      : | :||               ||| :     :|  | |  |
Db        263 ITFNGTDGTGRLVLLSKNAAATDFNVTGSLGGNLK---GIIEFNTVAVNG---------- 309

Qy        243 GAEWAEKADGNVRIVG-----------CAVCGESKKAQVKEQLARKIEWLPSSAQALGIR 291
                :    | |  ::|                  | | :  |:  | : :  || |:|
Db        310 --QLKANAGANAAVIGTNNGAGRAAGFVVSVDNGKVATIDGQVYAK-DMVIQSANAVGQV 366

Qy        292 NHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDA---LTDDGHYLGGTIMPGF 348
              | :||  : |:|      |:  |   |  |     |:| ::    || |:     |:|
Db        367 N-FRHIVDVGTD------GTTAFKTAASKV-----AITQNSNFGTTDFGNLAAQIIVPNT 414

Qy        349 HLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD---AVCGSIMMMHGRLKEKNGA 405
                :  :    | : :|                :||   |   ||  :|  :     | :||
Db        415 MTLNGNF---TGDASNPGNTAGVITFDANGTLASASADANVAVTNNITAI-----EASGA 466

Qy        406 GKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGG 455
              |    | ::|  ||::         | :  ::||  ||:| :|  |  ||
Db        467 GV---VQLSGTHAAELRLG-----NAGSVFKLADGTVINGKVNQTALVGG 508


RESULT 9
OMPB_RICJA
ID   OMPB_RICJA     STANDARD;      PRT;  1656 AA.
AC   O06653;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia japonica.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=35790;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YH;
RA   Uchiyama T.;
RT   "Sequencing of the gene encoding the protein rOmp B of Rickettsia
RT   japonica.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THE 120 KDA SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION (BY
CC       SIMILARITY).
CC   -!- FUNCTION: THE 32 KDA BETA PEPTIDE MAY SERVE AS A MEMBRANE ANCHOR
CC       (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB003681; BAA20138.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall.
FT   CHAIN         1   1338       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1339   1656       32 KDA BETA PEPTIDE.
FT   DOMAIN      528    533       POLY-GLY.
SQ   SEQUENCE   1656 AA;  168097 MW;  3132A69C9DD5999F CRC64;

  Query Match             5.4%;  Score 128;  DB 1;  Length 1656;
  Best Local Similarity   21.9%;  Pred. No. 0.3;
  Matches   94;  Conservative   59;  Mismatches  183;  Indels   94;  Gaps   18;

Qy         69 GKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAE 128
              |||: : | ||  |      || :: : |   :   | |        |: || :     :
Db        130 GKTLTITGQGITNVQSAATHNAQNIVAKFNGGAAIANND--------LSGLGTI-----D 176

Qy        129 EGFAPFLNEYETANRDHGKAVLLLRDGETVCEG-------------------TVKGV-- 166
               | |     :: ||    || |:| |   :  |                      |||  :
Db        177 FGAAASTLVFDLANPTTQKAPLILADNALIVNGANGTLNVTNGFIQVSDKSFATVKAINI 236

Qy        167 -DGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLE------------- 212
               ||:| :   |     :  ::|:    :::       : | :||
Db        237 GDGQGFMFNTNATN-----ANALNLQAGGTTINFNGTDGTGRLVLLSKNGAATDFNVTGS 291

Qy        213 -GGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVK 271
               ||| :     :|  | |  |     :  | |      :|  |  |  |  :: ||
Db        292 LGGNLK---GIIELNTVAINGQL-IANAGPANAVIGTN-NGAGRAAGFVVSVDNGKAATI 346

Qy        272 EQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVD 331
              :      : :  || | |  | :||  : | |      |:  |   | :|      |:| :
Db        347 DGQVYAKDMVIQSANANGQVN-FRHIVDVGID------GTTAFKTAASIV-----AITQN 394

Qy        332 A---LTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD-- 386
              :    || |:     :|    :  :     | : | |                :||   |
Db        395 SNFGTTDFGNLAAQVTVPDTMTLTGNF---TGDANNPGNTAGVITFAANGTLASASADAN 451

Qy        387 -AVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHG 445
               ||  :|  :     | :| |    | ::|   |::         | :  ::||  ||:|
Db        452 VAVTNNITAI-----EASGVGV---VQLSGTHTAELRLG-----NAGSVFKLADGTVING 498
```

```
Qy        446 LLNLIAAEGG 455
              :|     ||
Db        499 KVNQTVLVGG 508


RESULT 10
OMPB_RICRI
ID   OMPB_RICRI     STANDARD;      PRT;  1654 AA.
AC   Q53047;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein B precursor (168 kDa surface-layer protein)
DE   (Surface protein antigen) (Cell surface antigen 5) (Sca5) (rOmpB)
DE   (rOmp B) [Contains: 120 kDa surface-exposed protein (Surface protein
DE   antigen) (120 kDa outer membrane protein ompB); 32 kDa beta peptide].
GN   OMPB.
OS   Rickettsia rickettsii.
OC   Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC   Rickettsiaceae; Rickettsieae; Rickettsia.
OX   NCBI_TaxID=783;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R;
RX   MEDLINE=92167802; PubMed=1724278;
RA   Gilmore R.D. Jr., Cieplak W. Jr., Policastro P.F., Hackstadt T.;
RT   "The 120 kilodalton outer membrane protein (rOmp B) of Rickettsia
RT   rickettsii is encoded by an unusually long open reading frame:
RT   evidence for protein processing from a large precursor.";
RL   Mol. Microbiol. 5:2361-2370(1991).
RN   [2]
RP   SEQUENCE OF 279-1654 FROM N.A.
RC   STRAIN=R;
RX   MEDLINE=90136087; PubMed=2515418;
RA   Gilmore R.D. Jr., Joste N., McDonald G.A.;
RT   "Cloning, expression and sequence analysis of the gene encoding the
RT   120 kD surface-exposed protein of Rickettsia rickettsii.";
RL   Mol. Microbiol. 3:1579-1586(1989).
CC   -!- FUNCTION: THE 120 kDa SURFACE-EXPOSED PROTEIN IS A MAJOR
CC       STRUCTURAL PROTEIN WHICH MAY PLAY A ROLE AS A RICKETTSIAL
CC       VIRULENCE FACTOR AND/OR IMMUNOGEN DURING INFECTION.
CC   -!- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
CC       LAYER WITH HEXAGONAL SYMMETRY.
CC   -!- SIMILARITY: BELONGS TO THE RICKETTSIAE OMPA/OMPB FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X16353; CAA34403.1; -.
DR   InterPro; IPR003858; rOmpA_rOmpB.
DR   Pfam; PF02708; rOmpA_rOmpB; 1.
KW   Antigen; S-layer; Cell wall.
FT   CHAIN         1   1333       120 KDA SURFACE-EXPOSED PROTEIN.
FT   CHAIN      1334   1654       32 KDA BETA PEPTIDE.
FT   DOMAIN     1181   1188       POLY-THR.
SQ   SEQUENCE   1654 AA;  168184 MW;  D7AB70FB7087F618 CRC64;

  Query Match             4.9%;  Score 118;  DB 1;  Length 1654;
  Best Local Similarity   21.9%;  Pred. No. 1.6;
  Matches   98;  Conservative   60;  Mismatches  160;  Indels  130;  Gaps   21;

Qy         69 GKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAE 128
              |||: : | |:        :|| :|   |   :   | |        |  :| :    :
Db        130 GKTLTITGQGVTNAQAAATKNAQNVVVQFNNGAAIDNND--------LKGVGRI-----D 176

Qy        129 EGFAPFLNEYETANRDHGKAVLLLRDGETVCEG--------------------TVKGV-- 166
               |       :  ||    || |:| |   :  |                     |||  :
Db        177 FGAPASTLVFNLANPTTQKAPLILGDNAVIANGVNGTLNVTNGFIQVSNKSFATVKAINI 236

Qy        167 -DGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLE------------- 212
               ||:|::    |    |     ::|:    :::       : | :||
Db        237 ADGQGIIFNTDANNANT-----LNLQAGGTTINFTGTDGTGRLVLLSKHAAATNFNITGS 291

Qy        213 -GGNSR----LKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGES-K 266
               ||| :          |:    |  |:|     : :|       :|  |  |  |  :: |
Db        292 LGGNLKGVIEFNTVAVDGQLTANAGAAN----AVIGTN-----NGAGRAAGFVVSVDNGK 342

Qy        267 KAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 326
               | :  |:  | : :  || | |  | :||  : |:|      |:  |   |  |
Db        343 VATIDGQVYAK-DMVIQSANATGQVN-FRHIVDVGAD------GTTAFKTAASKV----- 389

Qy        327 AVTVD------------------ALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAG 367
               :| |                 |:|  |:: |    ||      :  | | : |
Db        390 TITQDSNFGNTDFGNLAAQIKVPNAITLTGNFTGDASNPG-----NTAGVITFDAN---- 440

Qy        368 KRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPP 427
                  |  :| |   : ||  :|  :     | :|||    | ::|  ||::
Db        441 ------GTLESASADANV-AVTNNITAI-----EASGAGV---VQLSGTHAAELRLG--- 482

Qy        428 AFLAENTVRVADNLVIHGLLNLIAAEGG 455
               | :  ::||  ||:| :|  |  ||
Db        483 --NAGSIFKLADGTVINGKVNQTALVGG 508


RESULT 11
RELA_STRCO
ID   RELA_STRCO     STANDARD;      PRT;   847 AA.
AC   P52560; P72401;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   GTP pyrophosphokinase (EC 2.7.6.5) (ATP:GTP 3'-pyrophosphotransferase)
DE   (PPGPP synthetase I) ((P)PPGPP synthetase).
GN   RELA OR SCO1513 OR SCL2.03C.
OS   Streptomyces coelicolor.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=96423179; PubMed=8825780;
RA   Chakraburtty R., White J., Takano E., Bibb M.J.;
RT   "Cloning, characterization and disruption of a (p)ppGpp synthetase
RT   gene (relA) of Streptomyces coelicolor A3(2).";
RL   Mol. Microbiol. 19:357-368(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / J802;
RX   MEDLINE=96209989; PubMed=8631867;
RA   Martenez-Costa O.H., Arias P., Romero N.M., Parro V., Mellado R.P.,
RA   Malpartida F.;
RT   "A relA/spoT homologous gene from Streptomyces coelicolor A3(2)
RT   controls antibiotic biosynthetic genes.";
RL   J. Biol. Chem. 271:10627-10634(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RX   MEDLINE=21996410; PubMed=12000953;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
```

```
CC   -!- FUNCTION: IN EUBACTERIA PPGPP (GUANOSINE 3'-DIPHOSPHATE 5-'
CC       DIPHOSPHATE) IS A MEDIATOR OF THE STRINGENT RESPONSE THAT
CC       COORDINATES A VARIETY OF CELLULAR ACTIVITIES IN RESPONSE TO
CC       CHANGES IN NUTRITIONAL ABUNDANCE. THIS ENZYME CATALYSES THE
CC       FORMATION OF PPPGPP WHICH IS THEN HYDROLYSED TO FORM PPGPP. PPPGPP
CC       COULD PLAY AN ESSENTIAL ROLE IN TRIGGERING ANTIBIOTIC PRODUCTION
CC       UNDER SOME NUTRITIONAL CONDITIONS.
CC   -!- CATALYTIC ACTIVITY: ATP + GTP = AMP + guanosine 3'-diphosphate 5'-
CC       triphosphate.
CC   -!- PATHWAY: FIRST STEP IN THE METABOLISM OF PPGPP.
CC   -!- INDUCTION: BY AMINO ACID STARVATION. ACTIVATION THROUGH THE
CC       BINDING OF UNCHARGED TRNA IN THE ACCEPTOR SITES OF TRANSLATING
CC       RIBOSOMES (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE RELA / SPOT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X87267; CAA60717.1; -.
DR   EMBL; X92520; CAA63297.1; -.
DR   EMBL; AL137778; CAB70915.1; -.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002819; HD.
DR   InterPro; IPR003607; ME_Pplase_HDc.
DR   InterPro; IPR004811; SpoT_relA.
DR   InterPro; IPR004095; TGS_dom.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF01966; HD; 1.
DR   Pfam; PF02824; TGS; 1.
DR   SMART; SM00471; HDc; 1.
DR   TIGRFAMs; TIGR00691; spoT_relA; 1.
KW   Transferase; Kinase; Complete proteome.
FT   CONFLICT    197    197       V -> L (IN REF. 2).
SQ   SEQUENCE   847 AA;  94181 MW;  544040B21A0D512E CRC64;

  Query Match             4.7%;  Score 111.5;  DB 1;  Length 847;
  Best Local Similarity   22.1%;  Pred. No. 2;
  Matches   77;  Conservative   47;  Mismatches  118;  Indels  107;  Gaps   14;

Qy         85 KEVENAASVQSLFQTASRRGNADAAV----LLETLLAELG--------AVLEQYAEEGFA 132
              :::| |  |   :    :| : |  :     : |:|||||         :|    |:
Db        125 RQIERAYQVAERWHRGQKRKSGDPYITHPLAVTTILAELGMDPATLMAGLLHDTVED--- 181

Qy        133 PFLNEY--ETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVSGEISL 190
                 ||  |   || |  | || || |  :    |          | |: | ||    :::
Db        182 ---TEYGLEDLRRDFGDVVTLLVDGVTKLDKVKFG---------EAAQAE-TVRKMVVAM 228

Qy        191 RPDNRSVSVPKRPD------SERFLLLEG------------------GNSRLKWAWVEN 225
              | | | | |  |      : |:|  |                    | : :||  :|:
Db        229 AKDPR-VLVIKLADRLHNMRTMRYLKREKQEKKARETLEIYAPLAHRLGMNTIKWE-LED 286

Qy        226 GTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSA 285
              || :    | ::  | || | | |  : :|   |  : : |::|  :  : : :   |
Db        287 LAFAILYPKMYDEIVRLVAERAPKRDEYLAVVTDEVQQDLRAARIKATVTGRPKHYYSVY 346

Qy        286 QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTD---------- 335
              | : :|                | |:    :|   |  | || : |
Db        347 QKMIVR------------------GRDFAEIYDLV---GIRVLVDTVRDCYAALGTVHA 384

Qy        336 -----DGHYLGGTIMPGFHLMKE------------SLAVRTANLNRPA 366
                  | :     || |:: :              | :|| :::| |
Db        385 RWNPVPGRFKDYIAMPKFNMYQSLHTTVIGPGGKPVELQIRTFDMHRRA 433


RESULT 12
FUMB_ECOLI
ID   FUMB_ECOLI     STANDARD;      PRT;   548 AA.
AC   P14407; P78139;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Fumarate hydratase class I, anaerobic (EC 4.2.1.2) (Fumarase).
GN   FUMB OR B4122.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=89255123; PubMed=2656658;
RA   Bell P.J., Andrews S.C., Sivak M.N., Guest J.R.;
RT   "Nucleotide sequence of the FNR-regulated fumarase gene (fumB) of
RT   Escherichia coli K-12.";
RL   J. Bacteriol. 171:3494-3503(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=95334362; PubMed=7610040;
RA   Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA   Blattner F.R.;
RT   "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT   region from 92.8 through 100 minutes.";
RL   Nucleic Acids Res. 23:2105-2119(1995).
CC   -!- FUNCTION: IT FUNCTIONS IN THE GENERATION OF FUMARATE FOR USE AS AN
CC       ANAEROBIC ELECTRON ACCEPTOR.
CC   -!- CATALYTIC ACTIVITY: (S)-malate = fumarate + H(2)O.
CC   -!- COFACTOR: BINDS 1 4FE-4S CLUSTER.
CC   -!- ENZYME REGULATION: SUBJECT TO ANAEROBIC REPRESSION.
CC   -!- PATHWAY: FERMENTATIVE PATHWAY THAT LEADS TO THE PRODUCTION OF
CC       SUCCINATE.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SIMILARITY: TO OTHER THERMOLABILE CLASS I FUMARASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27058; AAA23827.1; -.
DR   EMBL; U14003; AAA97022.1; -.
DR   EMBL; AE000485; AAC77083.1; -.
DR   PIR; B44511; B44511.
DR   EcoGene; EG10357; fumB.
DR   InterPro; IPR000362; Fumarate_lyase.
DR   InterPro; IPR004646; TtdA_fumA_fumB.
DR   InterPro; IPR004647; TtdB_fumA_fumB.
DR   Pfam; PF00206; lyase_1; 1.
DR   TIGRFAMs; TIGR00722; ttdA_fumA_fumB; 1.
DR   TIGRFAMs; TIGR00723; ttdB_fumA_fumB; 1.
DR   PROSITE; PS00163; FUMARATE_LYASES; 1.
KW   Lyase; Tricarboxylic acid cycle; Iron-sulfur; 4Fe-4S;
KW   Complete proteome.
FT   METAL       318    318       IRON-SULFUR (4FE-4S) (BY SIMILARITY).
FT   ACT_SITE    397    397       POTENTIAL.
FT   BINDING     463    463       CARBOXYL GROUP (POTENTIAL).
FT   CONFLICT     50     50       V -> L (IN REF. 2; AAC77083).
SQ   SEQUENCE   548 AA;  60091 MW;  336854D8FA5551AB CRC64;

  Query Match             4.7%;  Score 111;  DB 1;  Length 548;
  Best Local Similarity   21.1%;  Pred. No. 1.3;
  Matches   97;  Conservative   55;  Mismatches  166;  Indels  142;  Gaps   21;

Qy        113 ETLLAELGAVLEQYAEEGF--------------APFLNEYETANRDHGKAVLLLRDGET 157
                ||:|      |   |:: |              |  |:: | :  |   |:  ||: |
Db         37 ETILKVEPEALTLVAQQAFHDASFMLRPAHQKQVAAILHDPEASENDKYVALQFLRNSEI 96

Qy        158 VCEGTVKGVDGRGVLHLETAEG---------EQTVVSG------EISLRPD--------- 193
```

```
Db         97 AAKGVLPTCQDTGTAIIVGKKGQRVWTGGGDEETLSKGVYNTYIEDNLRYSQNAALDMYK 156

Qy        194 --NRSVSVPKR-------PDSERFL-LLEGGNS-----------------RLKWAWVEN- 225

Db        157 EVNTGTNLPAQIDLYAVDGDEYKFLCVAKGGGSANKTYLYQETKALLTPGKLKNFLVEKM 216

Qy        226 GTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGE-----------SKKAQVKEQL 274

Db        217 RTLGTAACPPYHIAFVIGGTSAETNLKTVKLASAHYYDELPTEGNEHGQAFRDVQLEQEL 276

Qy        275 ARKIEWLPSSAQALGIRNHYRHP------EEHGSDRWFNALGSRRFSRNACVV---VSCG 325

Db        277 LEEAQKLGLGAQFGG--KYFAHDIRVIRLPRHGA--------------SCPVGMGVSCS 319

Qy        326 TAVTVDA-LTDDG-------HYLGGTIMPGFHLMKESLAVRTANLNRPAGK------RYP 371

Db        320 ADRNIKAKINREGIWIEKLEHNPGQYIPQELRQAGEGEAVK-VDLNRPMKEILAQLSQYP 378

Qy        372 FPTTTGNAVASGMMDAVCGSIM----MMHGRLKEKNGAGKPV-----DVIITGGGAAKVA 422

o         379 VSTRL----------SLTGTIIVGRDIAHAKLKELIDAGKELPQYIKDHPIYYAGPAKTP 428

Qy        423 EALPPAFLAENTVRVADNLV----IHGLLNLIAAEGGESE 458

Db        429 AGYPSGSLGPTTAGRMDSYVDLLQSHGGSMIMLAKGNRSQ 468
```

```
RESULT 13
SAHH_SULSO
ID   SAHH_SULSO     STANDARD;      PRT;   417 AA.
AC   P50252; Q9UXE3;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE   hydrolase) (AdoHcyase).
GN   AHCY OR SSO0755.
OS   Sulfolobus solfataricus.
OC   Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC   Sulfolobus.
OX   NCBI_TaxID=2287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=97080494; PubMed=8921839;
RA   Porcelli M., Cacciapuoti G., Fusco S., Bertoldo C., de Rosa M.,
RA   Zappia V.;
RT   "Cloning and sequencing of the gene coding for S-adenosylhomocysteine
RT   hydrolase in the thermophilic archaeon Sulfolobus solfataricus.";
RL   Gene 177:17-22(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=20165948; PubMed=10701121;
RA   Charlebois R.L., Singh R.K., Chan-Weiher C.C.-Y., Allard G., Chow C.,
RA   Confalonieri F., Curtis B., Duguet M., Erauso G., Faguy D.,
RA   Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA   Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
RA   St Jean A., van der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
RA   Ragan M.A., Sensen C.W.;
RT   "Gene content and organization of a 281-kbp contig from the genome of
RT   the extremely thermophilic archaeon, Sulfolobus solfataricus P2.";
RL   Genome 43:116-136(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 35092 / DSM 1617 / P2;
RX   MEDLINE=21332296; PubMed=11427726;
RA   She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA   Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA   De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA   Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA   Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA   Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA   Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT   "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC   -!- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = adenosine
CC       + L-homocysteine.
CC   -!- COFACTOR: NAD (By similarity).
CC   -!- PATHWAY: Activated methyl cycle.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ADENOSYLHOMOCYSTEINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z50174; CAA90536.1; -.
DR   EMBL; Y18930; CAB57547.1; ALT_INIT.
DR   EMBL; AE006699; AAK41050.1; ALT_INIT.
DR   HSSP; P10760; 1B3R.
DR   InterPro; IPR000043; Ado_hcyase.
DR   Pfam; PF00670; AdoHcyase; 1.
DR   TIGRFAMs; TIGR00936; ahcY; 1.
DR   PROSITE; PS00738; ADOHCYASE_1; 1.
DR   PROSITE; PS00739; ADOHCYASE_2; 1.
KW   Hydrolase; NAD; One-carbon metabolism; Complete proteome.
FT   NP_BIND     207    238       NAD (POTENTIAL).
SQ   SEQUENCE   417 AA;  45965 MW;  7CD55A391ED7B4C9 CRC64;

  Query Match             4.4%;  Score 105;  DB 1;  Length 417;
  Best Local Similarity   22.7%;  Pred. No. 2.4;
  Matches   90;  Conservative   51;  Mismatches  152;  Indels  104;  Gaps   19;

Qy         76 GIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEEGFAPFL 135

Db         43 GINISAVLHVTKETAALVKTL-----KIGGANVALAGSNPLSTQDDVAAALVEEGISVFA 97

Qy        136 --NEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLH------LETAEGEQTVVSGE 187

Db         98 WKGENETEYYSNIESIVKIHEPNIVMD---DGADLHAYIHEKVSSKLDIYGGTEETTTGV 154

Qy        188 ISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWA 247

Db        155 IRLKAMEKD-GVLKYP------LVAVNNAYTKYLF--DNRYGTGQSA------------ 192

Qy        248 EKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFN 307

Db        193 --IDGILRATNILIAG-------KIAVVAGYGWV-----GRGIANRLR------------ 226

Qy        308 ALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAG 367

Db        227 GMGAR------VIVTEVDPIRALEA----------VMDGFDVMPIAEASKVGDI----- 264

Qy        368 KRYPFPTTTGNAVA---SGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEA 424

Db        265 ----FVTATGNTKAIRVEHMLNMKDGAILSNAGHFNVE------VDVKGLKETAVKVRNI 314

Qy        425 LP--PAFLAENTVRVADNLVIHG-LLNLIAAEGGESE 458

Db        315 RPYVDEYTLPNGKRV--YLLADGRLVNLAAAEGHPSE 349
```

```
RESULT 14
RELA_STRAT
ID   RELA_STRAT     STANDARD;      PRT;   841 AA.
AC   O85709;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   GTP pyrophosphokinase (EC 2.7.6.5) (ATP:GTP 3'-pyrophosphotransferase)
```

```
DE   (PPGPP synthetase I) ((P)PPGPP synthetase).
GN   RELA.
OS   Streptomyces antibioticus.
OC   Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1890;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=IMRU 3720;
RX   MEDLINE=99296594; PubMed=10368159;
RA   Hoyt S., Jones G.H.;
RT   "relA is required for actinomycin production in Streptomyces
RT   antibioticus.";
RL   J. Bacteriol. 181:3824-3829(1999).
CC   -!- FUNCTION: In eubacteria PPGPP (guanosine 3'-diphosphate 5-'
CC       diphosphate) is a mediator of the stringent response that
CC       coordinates a variety of cellular activities in response to
CC       changes in nutritional abundance. This enzyme catalyzes the
CC       formation of PPPGPP which is then hydrolysed to form PPGPP (By
CC       similarity). Is required for actinomycin production.
CC   -!- CATALYTIC ACTIVITY: ATP + GTP = AMP + guanosine 3'-diphosphate 5'-
CC       triphosphate.
CC   -!- PATHWAY: FIRST STEP IN THE METABOLISM OF PPGPP.
CC   -!- SIMILARITY: BELONGS TO THE RELA / SPOT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF072829; AAC26021.1; -.
DR   InterPro; IPR002912; ACT.
DR   InterPro; IPR002819; HD.
DR   InterPro; IPR003607; ME_Pplase_HDc.
DR   InterPro; IPR004811; SpoT_relA.
DR   InterPro; IPR004095; TGS_dom.
DR   Pfam; PF01842; ACT; 1.
DR   Pfam; PF01966; HD; 1.
DR   Pfam; PF02824; TGS; 1.
DR   SMART; SM00471; HDc; 1.
DR   TIGRFAMs; TIGR00691; spoT_relA; 1.
KW   Transferase; Kinase.
SQ   SEQUENCE   841 AA;  93671 MW;  632A037BA4EF4C94 CRC64;

  Query Match             4.4%;  Score 104;  DB 1;  Length 841;
  Best Local Similarity   21.4%;  Pred. No. 6.8;
  Matches   80;  Conservative   50;  Mismatches  129;  Indels  114;  Gaps   15;

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAV----LLETLL 116
              |:  |||    :   :       :::| |  |   :    :| : |  :     : |:|
Db        105 LLRIVRANDPKIETSTL-------RQIEKAYQVAERWHRGQKRKSGDPYITHPLAVTTIL 157

Qy        117 AELG--------AVLEQYAEEGFAPFLNEY--ETANRDHGKAVLLLRDGETVCEGTVKGV 166
              ||||        :|    |:       ||  :   || |  | || || |  :    |
Db        158 AELGMDPATLMAGLLHDSRED------TEYGLDDLRRDFGDVVGLLVDGVTKLDKVKFG- 210

Qy        167 DGRGVLHLETAEGEQTVVSGEISLRPDNRSVSVPKRPD------SERFLLLEG------- 213
                    | |: | ||    ::   | | | | |  |      : |:|  |
Db        211 --------EAAQAE-TVRKMVVAMAKDPR-VLVIKLADRLHNMRTMRYLKREKQEKKARE 260

Qy        214 -----------GNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAV 261
                        | : :||  :|:  || :    | ::  | || | | |  : ||   |
Db        261 TLEIYAPLAHRLGMNTIKWE-LEDLAFAILYPKMYDEIVRLVAERAPKRDEYLAIVTDEV 319

Qy        262 CGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 321
              : :  ::|  :  : :   |  | : :|                    | |:    :|
Db        320 QSDLRARRIKATVTGRPKHYYSVYQKIIVR------------------GRDFAEIYDLV 360

Qy        322 VSCGTAVTVDALTD--------------DGHYLGGTIMPGFHLMKE------------ 353
                  |  | || : |              | :     || |:: :
```

```
Db        361 ---GIRVLVDTVRDCYAALGTVHARWNPVPGRFKDYIAMPKFNMYQSLHTTVIGPNGKPV 417

Qy        354 SLAVRTANLNRPA 366
                | :|| :::| |
Db        418 ELQIRTFDMHRRA 430


RESULT 15
CH60_PANAN
ID   CH60_PANAN     STANDARD;      PRT;   541 AA.
AC   O66218;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN   GROL OR MOPA OR GROEL.
OS   Pantoea ananas (Erwinia uredovora).
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Pantoea.
OX   NCBI_TaxID=553;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JCM 6986;
RA   Harada H., Ishikawa H.;
RT   "Phylogenetical relationship based on groE genes among phenotypically
RT   related Enterobacter, Pantoea, Klebsiella, Serratia, and Erwinia
RT   species.";
RL   J. Gen. Appl. Microbiol. 43:355-361(1997).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS (BY SIMILARITY).
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB008151; BAA25235.1; -.
DR   HSSP; P06139; 1DK7.
DR   InterPro; IPR001844; Chaprnin_Cpn60.
DR   InterPro; IPR002423; Cpn60/TCP-1.
DR   Pfam; PF00118; cpn60_TCP1; 1.
DR   PRINTS; PR00298; CHAPERONIN60.
DR   PRINTS; PR00304; TCOMPLEXTCP1.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; 1.
KW   Chaperone; ATP-binding.
FT   NON_TER     541    541
SQ   SEQUENCE   541 AA;  56817 MW;  7111AA03A703E543 CRC64;

  Query Match             4.3%;  Score 103.5;  DB 1;  Length 541;
  Best Local Similarity   20.3%;  Pred. No. 4.2;
  Matches   77;  Conservative   54;  Mismatches  148;  Indels  101;  Gaps   14;

Qy         10 GWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI-----LIET 64
              |  |||     | |||       :   | : |||        :::   |:  |     ::|
Db        192 GMQFDR-----GYLSPY--FINKPETGAVELETPF-----ILLADKKISNIREMLPVLEA 239

Qy         65 VRAGGKTVAVVGIGINFVLPKEVENAA-------SVQSLFQTASRRGNA----------D 107
              |   || : ::          ::||  |       ::: : : |: :              |
Db        240 VAKAGKPLLIIA--------EDVEGEALATLVVNTMRGIVKVAAVKAPGFGDRRKAMLQD 291

Qy        108 AAVLL--ETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKG 165
               |:|     :  |:|  ||: | |                  | |:: :|  |: :|
Db        292 IAILTGGTVISEEIGMELEKAALEDLG-----------QAKRVVINKDTTTIIDGV--- 336

Qy        166 VDGRGVLHLETAEGEQTVVSG---EISLRPDNRSVSVPKRPDSERFLLLEGGNSRLK--- 219
                        ||:| : |   :|  : :  :    |    ||   | || : ||
```

```
Db        337 ------------GEETTIQGRVTQIRQQIEEATSDYDKEKLQERVAKLAGGVAVLKVGA 383

Qy        220 --------------------WAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGC 259
                                  | || |  |  | |  |  : |     :  | || |
Db        384 ATEVEMKEKKARVEDALHATRAAVEEGVVAGGGVALVRVAAQLTELRGQNEDQNVGIKVA 443

Qy        260 AVCGESKKAQVKEQLARKIEWLPSSAQA-LGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 318
                 ||   |:      :   : :: :|  |   :    ||:|:   |  |   : :|:|
Db        444 LRAMESPLRQIVSNAGEEPSVVANNVKAGDGNYGYNAQTEEYGNMIDFGILDPTKVTRSA 503

Qy        319 CVVVS--CGTAVTVDALTDD 336
                :   |  :| : :  |
Db        504 LQYAASVAGLMITTECMVTD 523

Search completed: June 24, 2003, 21:51:32
Job time : 9.00068 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 22:08:46 ; Search time 20.9726 Seconds
(without alignments)
2373.336 Million cell updates/sec

Title: US-09-813-453A-39
Perfect score: 2384
Sequence: 1 MGECLMFSFGWAFDRPQYEL.....LVIHGLLNLIAAEGGESEHA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2384	100.0	460	9	US-09-813-453A-39	Sequence 39, Appl
2	2223	93.2	592	9	US-09-813-453A-22	Sequence 22, Appl
3	2197	92.2	592	9	US-09-813-453A-43	Sequence 43, Appl
4	258.5	10.8	267	9	US-09-813-453A-15	Sequence 15, Appl
5	232.5	9.8	248	9	US-09-813-453A-20	Sequence 20, Appl
6	232	9.7	249	9	US-09-813-453A-61	Sequence 61, Appl
7	204	8.6	302	9	US-10-260-877-110	Sequence 110, App
8	199.5	8.4	242	9	US-09-813-453A-65	Sequence 65, Appl
9	185	7.8	260	9	US-09-813-453A-51	Sequence 51, Appl
10	179	7.5	265	9	US-09-813-453A-4	Sequence 4, Appli
11	177	7.4	249	9	US-09-813-453A-70	Sequence 70, Appl
12	169	7.1	250	9	US-09-813-453A-3	Sequence 3, Appli
13	168	7.0	262	9	US-09-813-453A-45	Sequence 45, Appl
14	166	7.0	244	9	US-09-813-453A-41	Sequence 41, Appl
15	161.5	6.8	257	9	US-09-813-453A-53	Sequence 53, Appl
16	153	6.4	219	9	US-09-813-453A-57	Sequence 57, Appl
17	153	6.4	288	9	US-09-738-626-4286	Sequence 4286, Ap
18	152.5	6.4	246	9	US-09-813-453A-9	Sequence 9, Appli
19	151	6.3	258	9	US-09-813-453A-2	Sequence 2, Appli
20	147.5	6.2	311	10	US-09-859-270-2	Sequence 2, Appli
21	146.5	6.1	272	9	US-09-712-363-276	Sequence 276, App
22	146.5	6.1	272	9	US-09-813-453A-5	Sequence 5, Appli
23	146.5	6.1	273	9	US-09-813-453A-10	Sequence 10, Appl
24	145	6.1	258	9	US-09-813-453A-49	Sequence 49, Appl
25	143	6.0	255	9	US-09-813-453A-7	Sequence 7, Appli
26	139	5.8	258	9	US-09-813-453A-6	Sequence 6, Appli
27	138.5	5.8	256	9	US-09-813-453A-55	Sequence 55, Appl
28	137	5.7	254	9	US-09-813-453A-47	Sequence 47, Appl
29	137	5.7	257	9	US-09-813-453A-13	Sequence 13, Appl
30	134	5.6	233	9	US-09-813-453A-17	Sequence 17, Appl
31	133.5	5.6	212	9	US-09-813-453A-59	Sequence 59, Appl
32	123.5	5.2	241	9	US-09-813-453A-63	Sequence 63, Appl
33	116	4.9	223	9	US-09-895-913A-74	Sequence 74, Appl
34	116	4.9	223	9	US-09-813-453A-14	Sequence 14, Appl
35	116	4.9	223	9	US-09-813-453A-67	Sequence 67, Appl
36	101.5	4.3	262	9	US-09-813-453A-8	Sequence 8, Appli
37	100	4.2	1981	9	US-09-928-457-38	Sequence 38, Appl
38	100	4.2	2015	9	US-10-066-551-1	Sequence 1, Appli
39	100	4.2	4613	9	US-09-860-846-31	Sequence 31, Appl
40	100	4.2	4613	9	US-09-988-384B-31	Sequence 31, Appl
41	100	4.2	4613	9	US-09-836-821-31	Sequence 31, Appl
42	100	4.2	4613	10	US-09-861-289-31	Sequence 31, Appl
43	100	4.2	11877	9	US-09-860-846-6	Sequence 6, Appli
44	100	4.2	11877	9	US-09-836-821-6	Sequence 6, Appli
45	100	4.2	11877	10	US-09-861-289-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-09-813-453A-39
; Sequence 39, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
; APPLICANT: Yocum, R. Rogers
; APPLICANT: Patterson, Thomas A.
; TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: ANTIBIOTICS
; FILE REFERENCE: OGZ-001
; CURRENT APPLICATION NUMBER: US/09/813,453A
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: US 60/227,860
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/667,569
; PRIOR FILING DATE: 2000-09-21
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-09-813-453A-39

Query Match 100.0%; Score 2384; DB 9; Length 460;
Best Local Similarity 100.0%; Pred. No. 8.9e-195;
Matches 460; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460
              |||||||||||||||||||||||||||||||||||||||||
Db        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEHA 460


RESULT 2
US-09-813-453A-22
  Sequence 22, Application US/09813453A
  Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-22

  Query Match           93.2%;  Score 2223;  DB 9;  Length 592;
  Best Local Similarity   93.7%;  Pred. No. 6.5e-181;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:||||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||| |||||||||||||| |||||:||||||||||||||||||||||||||||||
Db        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| ||||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 3
US-09-813-453A-43
; Sequence 43, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 43
;   LENGTH: 592
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-813-453A-43

  Query Match           92.2%;  Score 2197;  DB 9;  Length 592;
  Best Local Similarity   92.6%;  Pred. No. 1.1e-178;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||: ||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:||||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||||||||||||||||||| |||||:||||||||||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| ||||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||:||||:||||| | ||
Db        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 4
US-09-813-453A-15
; Sequence 15, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
```

```
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
;   LENGTH: 267
;   TYPE: PRT
;   ORGANISM: Bordetella pertussis
US-09-813-453A-15

  Query Match             10.8%;  Score 258.5;  DB 9;  Length 267;
  Best Local Similarity   31.8%;  Pred. No. 4.3e-14;
  Matches   84;  Conservative   43;  Mismatches  116;  Indels   21;  Gaps    8;

Qy        209 LLLEGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 263
              :|:: ||||||  | : :   |    ||      ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        264 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 317
              ::   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        318 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTG 377
              :| | ||| |:| :   |  : || |:||  :|: :||  ||:|    |    :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        378 NAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGG-------AAKVAEALPPAFL 430
              |:|||:  |  |:|:     |  :   |:  :: : |||       | ::      ||
Db        183 QAIASGIAAAQAGAIVRQW--LAGRQRYGQAPEIYVAGGGWPEVRQEAERLLAVTGAAFG 240

Qy        431 AENTVRVADNLVIHGLLNLIAAEG 454
              |      |: |: ||  | ||:|
Db        241 ATPQPTYLDSPVLDGLAAL-AAQG 263


RESULT 5
US-09-813-453A-20
; Sequence 20, Application US/09813453A
  Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-813-453A-20

  Query Match             9.8%;  Score 232.5;  DB 9;  Length 248;
  Best Local Similarity   27.2%;  Pred. No. 6.3e-12;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy        209 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 268
              | |: ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Qy        269 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 321
              :  : :||  :  | ||        |  |:|| |   :  | |||   : :   :: ||:|
Db         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Qy        322 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 374
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Qy        375 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 429
                | | |  :     | ::|: | ::|:       |   :: :||| |  | : |  |
Db        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

Qy        430 LAENTVRVADNLVIHGL 446
                    |:  :||  ||
Db        231 ------RIMPDLVFVGL 241


RESULT 6
US-09-813-453A-61
; Sequence 61, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 61
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas putida
US-09-813-453A-61

  Query Match             9.7%;  Score 232;  DB 9;  Length 249;
  Best Local Similarity   29.0%;  Pred. No. 7e-12;
  Matches   78;  Conservative   41;  Mismatches   90;  Indels   60;  Gaps   12;

Qy        209 LLLEGGNSRLKW-------AWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGC-- 259
              | |: ||| :||        | :| |       |   ::: |          :||: ||
Db          3 LELDCGNSFIKWRVIHVADAVIEGGGIVDSDQALVAEVAAL---------ASVRLTGCRI 53

Qy        260 ----------AVCGESKKA-QVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNA 308
                         |:|    :|  |: ::|  :       :  |:|| |   :  | |||  |
Db         54 VSVRSEEETDALCALIAQAFAVQAKVAHPVR------EMAGVRNGYDDYQRLGMDRWLAA 107

Qy        309 LGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGK 368
              ||:   :: ||:|:  |||   | :: || :||| | ||  ||:  |   | :
Db        108 LGAFHLAKGACLVIDLGTAAKADFVSADGEHLGGYICPGMPLMRSQLRTHTRRI------ 161

Qy        369 RYPFPTTTGNAVAS-----GMMDAV-CGSIMMMHG----RLKEKNGA-GKPVDVIITGGG 417
              ||     :   |::|       ::||  | ::|: |    :|::     |:   | :|||
Db        162 RYD-DASAERALSSLSPGRSTVEAVERGCVLMLQGFAYTQLEQARVLWGEEFTVFLTGGD 220

Qy        418 AAKVAEALPPAFLAENTVRVADNLVIHGL 446
              |  |  ||| |       ||  :||  ||
Db        221 APLVRAALPQA-------RVVPDLVFVGL 242


RESULT 7
US-10-260-877-110
; Sequence 110, Application US/10260877
; Publication No. US20030021813A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Abbott Laboratories
;  APPLICANT:  Chovan, Linda E.
;  APPLICANT:  Hessler, Paul E.
;  APPLICANT:  Reich, Karl A.
;  TITLE OF INVENTION: ESSENTIAL BACTERIA GENES AND GENOME
;  TITLE OF INVENTION:  SCANNING IN HAEMOPHILUS INFLUENZAE FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  'ESSENTIAL GENES'
;  FILE REFERENCE: 6565.US.P1
;  CURRENT APPLICATION NUMBER: US/10/260,877
;  CURRENT FILING DATE:  2002-09-30
;  PRIOR APPLICATION NUMBER: US/09/649,145
;  PRIOR FILING DATE: 2000-08-25
;  NUMBER OF SEQ ID NOS: 137
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 110
;   LENGTH: 302
;   TYPE: PRT
;   ORGANISM: H. influenzae
US-10-260-877-110

  Query Match             8.6%;  Score 204;  DB 9;  Length 302;
  Best Local Similarity   32.3%;  Pred. No. 2.2e-09;
  Matches   61;  Conservative   27;  Mismatches   89;  Indels   12;  Gaps    6;

Qy          5 LMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIET 64
              :|||| |||| |:  :  || |  ||          |  |:||||::    ||||||:|
Db        122 IMFSFYWAFD-PKKSIEGLSLVIGLAIAEV-----LNVQVKWPNDILFDERKLGGILVEI 175

Qy         65 V-RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVL 123
                |    |:|||||  | |: | :   |   :      : :   ||  |:  |   |
Db        176 ANHKNGMLNLVIGIGINVSLSKQTEIS---QPYAEVCEIDPDVERQTLLPKLIQHLYTRL 232

Qy        124 EQYAEEGF-APFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQT 182
               : : |    |   ::: |      : :| : : |  |  :|:| || | :       |
Db        233 NIFEQNGIDEEFQQAWQSYNAFSNSEINVLTE-QGVISGIEQGIDERGYLKVLCGNKIQM 291

Qy        183 VVSGEISLR 191
                ||:|||
Db        292 FNGGEVSLR 300


RESULT 8
US-09-813-453A-65
; Sequence 65, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
   TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
   TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
;   LENGTH: 242
;   TYPE: PRT
;   ORGANISM: Xylessa fastidiosa
US-09-813-453A-65

  Query Match             8.4%;  Score 199.5;  DB 9;  Length 242;
  Best Local Similarity   30.3%;  Pred. No. 4e-09;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy        210 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 269
              | : |||| | | :  |    |   ||   :       |   | |  :  :|   :
Db          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63

Qy        270 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 326
              | | |      :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121

Qy        327 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 386
              |:|:| |  :| :||| |      ||:::|  |   |    |    |  | :|: ||
Db        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181

Qy        387 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 442
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228

Qy        443 IHGL 446
              : ||
Db        229 LDGL 232


RESULT 9
US-09-813-453A-51
; Sequence 51, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51
;   LENGTH: 260
;   TYPE: PRT
;   ORGANISM: Caulobacter crescentus
US-09-813-453A-51

  Query Match             7.8%;  Score 185;  DB 9;  Length 260;
  Best Local Similarity   34.1%;  Pred. No. 7.5e-08;
  Matches   58;  Conservative   23;  Mismatches   69;  Indels   20;  Gaps    6;

Qy        288 LGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG 347
              |||       | | |:||  ||:|:        ||:  ||| | | :  || : || | ||
Db         94 LGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPG 153

Qy        348 FHLMKESLAVRTANL-----NRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHG---RL 399
               :|  ::|    | |       |||| |    | | :|: ||:     | | :: |   |:
Db        154 INLSMQALHEAAAKLPRIAIQRPAGNRI-VGTDTVSAMQSGVF---WGYISLIEGLVARI 209

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNL 449
              | :   |:|: || ||| |:    |      ::      :| | ||| :
Db        210 KAER--GEPMTVIATGGVASLFEGA------TDSIDHFDSDLTIRGLLEI 251


RESULT 10
US-09-813-453A-4
; Sequence 4, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
```

```
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 265
;   TYPE: PRT
;   ORGANISM: Streptomyces coelicolor
US-09-813-453A-4

  Query Match             7.5%;  Score 179;  DB 9;  Length 265;
  Best Local Similarity   27.7%;  Pred. No. 2.5e-07;
  Matches   74;  Conservative   35;  Mismatches  106;  Indels   52;  Gaps   10;

Qy        178 EGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYR 237
              :||  |    ||   |:|      |   | :||:|                :|  |
Db         18 DGEDIVEHWRIS--TDSR------RTADELAVLLQG---------------LMGMHPL- 52

Qy        238 DLSPLGAEWAEKADGNVRIVGCAVCG--ESKKAQVKEQLARKIEWLPSSAQALGIRNHY- 294
              || |  :  |      | |:|   |   :::|   |    :|:     |::
Db         53 ----LGDELGDGID------GIAICATVPSVLHELREVTRRYYGDVPAVLVEPGVKTGVP 102

Qy        295 ---RHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLM 351
                 ||:| |:|| ||: :        :||  ||| | ||::  | |:|| | ||  :
Db        103 ILTDHPKEVGADRIINAVAAVELYGGPAIVVDFGTATTFDAVSARGEYIGGVIAPGIEIS 162

Qy        352 KESLAVRTANLNR--PAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRL-KEKNGAGKP 408
               |:| |: | | :   | |      |  |: ||::    | :  :  |: :|
Db        163 VEALGVKGAQLRKIEVARPRSVIGKNTVEAMQSGIVYGFAGQVDGVVNRMARELADDPDD 222

Qy        409 VDVIITGGGAAKVAEALPPAFLAENTV 435
              | || ||.|       | |  | |::|
Db        223 VTVIATGG--------LAPMVLGESSV 241


RESULT 11
US-09-813-453A-70
; Sequence 70, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
   PRIOR APPLICATION NUMBER: US 60/227,860
   PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 70
;   LENGTH: 249
;   TYPE: PRT
;   ORGANISM: Pseudomonas syringae
US-09-813-453A-70

  Query Match             7.4%;  Score 177;  DB 9;  Length 249;
  Best Local Similarity   28.1%;  Pred. No. 3.4e-07;
  Matches   72;  Conservative   31;  Mismatches  119;  Indels   34;  Gaps    8;

Qy        209 LLLEGGNSRLKWAWVENGTFATV-GSAPYRDLSPL-------GAEWAEKADGNVRIVGCA 260
              | |: ||| :||  :        | |      | : |       || :::      |:|  :
Db          3 LELDCGNSFIKWRIITKSCSTLVSGGVVDSDTALLECLGNLSGAAFSD-----CRLV--S 55

Qy        261 VCGESKKAQVKEQLARKIEWLPSSA----QALGIRNHYRHPEEHGSDRWFNALGSRRFSR 316
              |    : |::   ||       |  |    :  |: | |      | |||   :|:    :
Db         56 VRSAEETAKLVCALADTFSISPVCAAPAPELAGVINGYDDFARLGLDRWLAFVGAYHLVK 115

Qy        317 NACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN------RPAGKRY 370
               ||:|:  |||:| | :   | :||| | ||  ||:  |   |  :        |  |

Db        116 GACLVIDLGTAITSDFVEASGKHLGGFICPGMPLMRNQLRTHTRRIRYDDAEAEKALVRL 175

Qy        371 PFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFL 430
                  |  ||  |    : |  |      :|    |    : :||| |  ||: || |
Db        176 VPGRATAEAVERGCSLMLRGFAMTQIEIAREY--WGDDFAIFVTGGDAVLVADVLPGA-- 231

Qy        431 AENTVRVADNLVIHGL 446
                   |:  :||  ||
Db        232 -----RIVPDLVFVGL 242


RESULT 12
US-09-813-453A-3
; Sequence 3, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Clostridium acetobutylicum
US-09-813-453A-3

  Query Match             7.1%;  Score 169;  DB 9;  Length 250;
  Best Local Similarity   30.0%;  Pred. No. 1.6e-06;
  Matches   42;  Conservative   29;  Mismatches   59;  Indels   10;  Gaps    3;

Qy        289 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 348
              ||   | :|:| |:||  ||: :    : : :::  ||| |  |: ::| |||| | ||
Db        112 GINIKYDNPKEVGADRIVNAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGI 171

Qy        349 HLMKESLAVRTANLNR-----PAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEK- 402
                 :|:|  : | | |     ||   |     | ::: ||::      :  :  :|||
Db        172 KVSSEALFEKAAKLPRVELIKPA---YAICKNTISSIQSGIVYRYLRQVKYLFEKLKENL 228

Qy        403 -NGAGKPVDVIITGGGAAKV 421
               :|     :::  || ||:
Db        229 PDGRRTRTSLVLATGGLAKL 248


RESULT 13
US-09-813-453A-45
; Sequence 45, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
;   LENGTH: 262
;   TYPE: PRT
```

```
;   ORGANISM: Bacillus anthracis
US-09-813-453A-45

  Query Match             7.0%;  Score 168;  DB 9;  Length 262;
  Best Local Similarity   30.6%;  Pred. No. 2.1e-06;
  Matches   52;  Conservative   26;  Mismatches   68;  Indels   24;  Gaps    5;

Qy        289 GIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGF 348
              |:   | :| | |:||  ||:       :  ::|  ||| |   : :: ||:|| | ||
Db         95 GLNIKYENPREVGADRIVNAVAGIHLYGSPLIIVDFGTATTYCYINEEKHYMGGVITPGI 154

Qy        349 HLMKESLAVRTANLNR--------PAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLK 400
               :  |:|  | | | |          ||       | :|: ||::    | :  :  |:|
Db        155 MISAEALYSRAAKLPRIEITKPSSVVGK------NTVSAMQSGILYGYVGQVEGIVKRMK 208

Qy        401 EKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADN-LVIHGLLNL 449
              |:  | :   || ||| |  ::|         | : | |  | : ||  |
Db        209 EE--AKQEPKVIATGGLAKLISE-------ESNVIDVVDPFLTLKGLYML 249


RESULT 14
US-09-813-453A-41
; Sequence 41, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 41
;   LENGTH: 244
;   TYPE: PRT
;   ORGANISM: Porphyromonas gingivalis
US-09-813-453A-41

  Query Match             7.0%;  Score 166;  DB 9;  Length 244;
  Best Local Similarity   25.2%;  Pred. No. 2.9e-06;
  Matches   67;  Conservative   44;  Mismatches  101;  Indels   54;  Gaps    9;

Qy        209 LLLEGGNSRLKWAWVENGTFATVGSAPYR---DLSPLGAEWA------------EKADG 252
              |::: |||  | |:| | :  ::   | :    || | |                |:|:
Db          5 LIVDQGNSACKVAFVRNNSIESISFLPGKAGQALSHLVAPHRFDKAIYSSVGLPDEEAEA 64

Qy        253 NVRIVGCA--VCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG 310
               ||    |  : |      :: |  |:                        |:||    :|
Db         65 IVRSCAAASLMMGTETPVPLRLQYDRR-------------------TLGADRLAAVVG 103

Qy        311 SRRFSRNA-CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLN--RPAG 367
              :     |   :|:  |||:| : :: :| |||| | || ||  ::| : |  |    |:|
Db        104 AHSLYPNTELLVIDAGTAITYERVSAEGIYLGGNISPGLHLRFKALHLFTGRLPLIDPSG 163

Qy        368 ---KRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEA 424
                 |   : ::|  |: :|::  : | |      |  | |      ||:||| |  :|
Db        164 ISPKIAEYGSSTEEAITAGVIHGLAGEIDRYIDDLHAKEGRSA---VILTGGDANYLARI 220

Qy        425 LPPAFLAENTVRVADNLVIHGLLNLI 450
              :    |      :  :||: ||  ::
Db        221 IRSGIL------IHPDLVLLGLNRIL 240


RESULT 15
US-09-813-453A-53
; Sequence 53, Application US/09813453A
; Patent No. US20020168681A1
; GENERAL INFORMATION:
;  APPLICANT: Yocum, R. Rogers
;  APPLICANT:  Patterson, Thomas A.
;  TITLE OF INVENTION: MICROORGANISMS AND ASSAYS FOR THE IDENTIFICATION OF
;  TITLE OF INVENTION:  ANTIBIOTICS
;  FILE REFERENCE: OGZ-001
;  CURRENT APPLICATION NUMBER: US/09/813,453A
;  CURRENT FILING DATE:  2001-03-20
;  PRIOR APPLICATION NUMBER: US 60/227,860
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: 09/667,569
;  PRIOR FILING DATE: 2000-09-21
;  NUMBER OF SEQ ID NOS: 77
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 53
;   LENGTH: 257
;   TYPE: PRT
;   ORGANISM: Chlorobium tepidum
US-09-813-453A-53

  Query Match             6.8%;  Score 161.5;  DB 9;  Length 257;
  Best Local Similarity   24.5%;  Pred. No. 7.4e-06;
  Matches   68;  Conservative   33;  Mismatches   99;  Indels   77;  Gaps    9;

Qy        209 LLLEGGNSRLKWAWV---ENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGES 265
              |::: ||:    |     |  :  :| || : | | :   :   |  :      |:|
Db          3 LVVDIGNTSTTLAIFTGDEEPSVESVPSALFADSSTMREVFGNMARKHGEPQAIAICS-- 60

Qy        266 KKAQVKEQLARKIEWLPSSAQALG----------------------IRNHYRHPEEHGSD 303
                              :  || |:|                    |  |  |   |:|
Db         61 --------------VVPSATAVGSALLESLFSVPVLTICCKLRFPFRLDYATPHTFGAD 105

Qy        304 RWFNALGSRR-FSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANL 362
              |      ||  ||    : |  |||:| | |   |:| || ||||  :|  :|  ||| |
Db        106 RLALCAWSRHLFSEKPVIAVDIGTAITFDVLDTVGNYRGGLIMPGIDMMAGALHSRTAQL 165

Qy        363 -----NRPAGKRYPFPTTTGNAVAS----------GMMDAVCGSIMMMHGRLKEKNGAG 406
                   :||         :|   : |          |::||: | ::   |
Db        166 PQVRIDRPESL---LGRSTTECIKSGVFWGVVKQIGGLVDAIRGDLVRDFGE-------- 214

Qy        407 KPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVI 443
                |:||:||| :  :   : |       | | | | :
Db        215 STVEVIVTGGNSRIIVPEIGP-------VSVIDELAV 244


Search completed: June 24, 2003, 22:24:07
Job time : 21.9726 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:27:56 ; Search time 13.438 Seconds
(without alignments)
1007.182 Million cell updates/sec

Title: US-09-813-453A-39
Perfect score: 2384
Sequence: 1 MGECLMFSFGWAFDRPQYEL.....LVIHGLLNLIAAEGGESEHA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	147.5	6.2	311	4	US-09-175-172-2	Sequence 2, Appli
2	125.5	5.3	5588	4	US-09-036-987A-6	Sequence 6, Appli
3	125.5	5.3	5588	4	US-09-370-700-6	Sequence 6, Appli
4	125	5.2	323	4	US-09-237-119-2	Sequence 2, Appli
5	115	4.8	328	4	US-09-134-001C-3229	Sequence 3229, Ap
6	110	4.6	1611	2	US-08-804-227C-5	Sequence 5, Appli
7	101.5	4.3	1271	1	US-08-095-734-2	Sequence 2, Appli
8	101.5	4.3	1271	2	US-08-444-623-2	Sequence 2, Appli
9	101.5	4.3	1271	3	US-08-471-869-2	Sequence 2, Appli
10	101.5	4.3	1271	4	US-09-342-563-2	Sequence 2, Appli
11	101.5	4.3	1271	5	PCT-US94-08267-2	Sequence 2, Appli
12	100.5	4.2	4472	2	US-08-804-227C-2	Sequence 2, Appli
13	100	4.2	4613	4	US-09-105-537-31	Sequence 31, Appl
14	100	4.2	11877	4	US-09-105-537-6	Sequence 6, Appli
15	98.5	4.1	357	4	US-09-605-785-826	Sequence 826, App
16	98.5	4.1	1228	4	US-09-605-785-537	Sequence 537, App
17	98.5	4.1	1228	4	US-09-439-313-537	Sequence 537, App
18	98.5	4.1	1261	4	US-09-605-785-538	Sequence 538, App
19	98.5	4.1	1261	4	US-09-439-313-538	Sequence 538, App
20	98	4.1	447	1	US-08-476-008-67	Sequence 67, Appl
21	98	4.1	447	1	US-08-306-063-67	Sequence 67, Appl
22	98	4.1	447	1	US-08-833-485-67	Sequence 67, Appl
23	98	4.1	447	4	US-09-137-440-67	Sequence 67, Appl
24	98	4.1	454	4	US-09-319-892-2	Sequence 2, Appli
25	98	4.1	454	4	US-09-418-817-8	Sequence 8, Appli
26	97	4.1	1222	2	US-08-682-517-15	Sequence 15, Appl
27	97	4.1	1252	2	US-08-682-517-9	Sequence 9, Appli
28	97	4.1	4551	3	US-09-320-878-1	Sequence 1, Appli
29	94.5	4.0	5087	4	US-09-144-085-1	Sequence 1, Appli
30	94	3.9	609	4	US-08-969-683A-67	Sequence 67, Appl
31	92.5	3.9	1891	2	US-08-804-227C-12	Sequence 12, Appl
32	92.5	3.9	1891	2	US-08-804-198-6	Sequence 6, Appli
33	92.5	3.9	6095	4	US-09-144-085-2	Sequence 2, Appli
34	92	3.9	915	4	US-09-206-942-35	Sequence 35, Appl
35	92	3.9	947	4	US-09-228-986-73	Sequence 73, Appl
36	92	3.9	1222	4	US-09-206-942-37	Sequence 37, Appl
37	92	3.9	1228	4	US-09-206-942-34	Sequence 34, Appl
38	92	3.9	1489	6	5183745-2	Patent No. 5183745
39	91.5	3.8	304	4	US-09-134-001C-4578	Sequence 4578, Ap
40	90	3.8	444	6	5310667-5	Patent No. 5310667
41	90	3.8	3729	2	US-08-804-227C-4	Sequence 4, Appli
42	89	3.7	915	1	US-08-487-890A-96	Sequence 96, Appl
43	89	3.7	915	2	US-08-363-124A-2	Sequence 2, Appli
44	89	3.7	915	2	US-08-478-435-96	Sequence 96, Appl
45	89	3.7	915	2	US-08-337-483-96	Sequence 96, Appl

ALIGNMENTS

RESULT 1
US-09-175-172-2
; Sequence 2, Application US/09175172
; Patent No. 6255075
; GENERAL INFORMATION:
; APPLICANT: Huang, Jianzhong
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: birA
; FILE REFERENCE: GM10172
; CURRENT APPLICATION NUMBER: US/09/175,172
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-175-172-2

Query Match 6.2%; Score 147.5; DB 4; Length 311;
Best Local Similarity 31.5%; Pred. No. 2.3e-07;
Matches 58; Conservative 33; Mismatches 70; Indels 23; Gaps 9;

```
Qy         16 PQYELGSLSPVAALACRRALGCLGL-ETQIKWPNDLVVGRDKLGGILIE---TVRAGGKT 71
              |  :| | : : | |  :|:  | | :  ||| ||: :   |:|||| |   :|  | |
Db        138 PYDKLPSYTLLVAGAVYKAIKNLTLIDVDIKWVNDIYLNNHKIGGILTEAMTSVETGLVT 197

Qy         72 VAVVGIGINFVL---PKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL-GAVLEQYA 127
                ::|:|||| :   |:|:: :|:  |||:       | | :    |: |:  |  | |
Db        198 DIIIGVGINFTIKDFPQELKEKAA--SLFK-------ATAPITRNELIIEIWRAFFETPA 248

Qy        128 EEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQT-VVSG 186
              ||    | |:  :  || |    : :   :|  | :   | | ::   |::  : ||
Db        249 EE----LLYLYKKQSFILGKEVTFTLEQKDY-KGLAKDISENGKLLVQCDNGKEIWLNSG 303

Qy        187 EISL 190
              ||||
Db        304 EISL 307
```

RESULT 2
US-09-036-987A-6
; Sequence 6, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.

```
;    APPLICANT:  Treadway, Patti J.
;    APPLICANT:  Turner, Jan R.
;    APPLICANT:  Waldron, Clive
;    TITLE OF INVENTION:  Biosynthetic Genes For Spinosyn Insecticide
;    TITLE OF INVENTION:  Production
;    NUMBER OF SEQUENCES:  39
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dow AgroSciences LLC Patent Department
;      STREET:  9330 Zionsville Road
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  USA
;      ZIP:  46268
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/036,987A
;      FILING DATE:  09-MAR-1998
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Stuart, Donald R
;      REGISTRATION NUMBER:  28,479
;      REFERENCE/DOCKET NUMBER:  50,608
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317)337-4816
;      TELEFAX:  (317)337-4847
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5588 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-036-987A-6

  Query Match           5.3%;  Score 125.5;  DB 4;  Length 5588;
  Best Local Similarity   18.8%;  Pred. No. 0.0066;
  Matches  119;  Conservative   69;  Mismatches  198;  Indels  247;  Gaps   24;

Qy         11 WAFDRPQYELGSLSPVAALACRRAL----------------GCLGLETQIKWPNDLVVG- 53
Db        907 W-LESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQ-PWLADHVVGG 964

Qy         54 ----------------RDKLGGILIETV------------------------RAGGKTVA 73
Db        965 AILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVR 1024

Qy         74 V----------------------------VGIGINFVLPKEVENAASVQSLFQTASRRG- 104
Db       1025 VHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPLDSFYARAAERGF 1084

Qy        105 ----------------------------NADAA------VLLETLLAELGAVLEQYAEE 129
Db       1085 DFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQALGAA-EEDPDE 1143

Qy        130 GFAPFLNEYET--ANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVSGE 187
Db       1144 GWLPFAWQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSID------SLVLRQ 1197

Qy        188 ISLRPDNRSVSVPKRPDSERFLLLE------GGNSRLKWAWVENGTFA-TVGSAPYRDLS 240
Db       1198 IS----DKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALGTELH 1253

Qy        241 PLGAEWAEKAD--GNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPE 298
Db       1254 P---DLTGLADPPPDVVVVPCG----------------------ASRQDLDVASEARAAT 1288

Qy        299 EHGSDRWFNALGSRRFSRNACVVVSCGTA------------------------------ 327
Db       1289 QRMLDLIQDWLAAARFAGSRLVVVTCGAASTGPAEGVSDLVHAASWGLLRSAQSENPDRF 1348

Qy        328 --VTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP--------AGKRYPFPTTTG 377
Db       1349 VLVDVDGTAESWRALAAAVRSG----EPQLALRAGEVRVPRLARCVAAEDSRIPVPGADG 1404

Qy        378 NAVASGMMDAVCGSI------------MMMHGRLKEKNGAGKP-----VDVIITGGGAAK 420
Db       1405 TVLISGGTGLLGGLVARHLVAERGVRRLVLAGR----RGWSAPGVTDLVDELVGLGAAVE 1460

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAE 453
Db       1461 VA-----------SCDVGDRAQLDRLLTTISAE 1482


RESULT 3
US-09-370-700-6
; Sequence 6, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
;  APPLICANT: Baltz, Richard H
;  APPLICANT: Broughton, Mary C
;  APPLICANT: Crawford, Kathryn P
;  APPLICANT: Madduri, Krishnamurthy
;  APPLICANT: Treadway, Patti J
;  APPLICANT: Turner, Jan R
;  APPLICANT: Waldron, Clive
;  TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
;  FILE REFERENCE: 50489 DIV1
;  CURRENT APPLICATION NUMBER: US/09/370,700
;  CURRENT FILING DATE: 1999-08-09
;  EARLIER APPLICATION NUMBER: US 09/36987
;  EARLIER FILING DATE: 1998-03-09
;  NUMBER OF SEQ ID NOS: 39
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 5588
;   TYPE: PRT
;   ORGANISM: Saccharopolyspora spinosa
US-09-370-700-6

  Query Match           5.3%;  Score 125.5;  DB 4;  Length 5588;
  Best Local Similarity   18.8%;  Pred. No. 0.0066;
  Matches  119;  Conservative   69;  Mismatches  198;  Indels  247;  Gaps   24;

Qy         11 WAFDRPQYELGSLSPVAALACRRAL----------------GCLGLETQIKWPNDLVVG- 53
Db        907 W-LESAEYDAGDLGSVGLLSAEHPLLGAAVTLADAGGFLLTGKLSVKTQ-PWLADHVVGG 964

Qy         54 ----------------RDKLGGILIETV------------------------RAGGKTVA 73
Db        965 AILLPGTAFVEMLIRAADQVGCDLIEELSLTTPLVLPATGAVQVQIAVGGPDEAGRRSVR 1024

Qy         74 V----------------------------VGIGINFVLPKEVENAASVQSLFQTASRRG- 104
Db       1025 VHSCRDDAVPQDSWTCHATGTLTSSDHQDAGQGPDGIWPPNDAVAVPLDSFYARAAERGF 1084

Qy        105 ----------------------------NADAA------VLLETLLAELGAVLEQYAEE 129
Db       1085 DFGPAFQGLQAAWKRGDEIFAEVGLPTAHREDAGRFGIHPALLDAALQALGAA-EEDPDE 1143

Qy        130 GFAPFLNEYET--ANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVSGE 187
Db       1144 GWLPFAWQGVSLKATGALSLRVHLVPAGANAVSVFTTDTTGQAVLSID------SLVLRQ 1197

Qy        188 ISLRPDNRSVSVPKRPDSERFLLLE------GGNSRLKWAWVENGTFA-TVGSAPYRDLS 240
Db       1198 IS----DKQLAAARAMEHESLFRVDWKRISPGAAKPVSWAVIGNDELARACGSALGTELH 1253

Qy        241 PLGAEWAEKAD--GNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPE 298
Db       1254 P---DLTGLADPPPDVVVVPCG----------------------ASRQDLDVASEARAAT 1288

Qy        299 EHGSDRWFNALGSRRFSRNACVVVSCGTA------------------------------ 327
```

```
Db       1289 QRMLDLIQDWLAAARFAGSRLVVVTCGAASTGPAEGVSDLVHAASWGLLRSAQSENPDRF 1348

Qy        328 --VTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP--------AGKRYPFPTTTG 377
Db       1349 VLVDVDGTAESWRALAAAVRSG----EPQLALRAGEVRVPRLARCVAAEDSRIPVPGADG 1404

Qy        378 NAVASGMMDAVCGSI------------MMMHGRLKEKNGAGKP-----VDVIITGGGAAK 420
Db       1405 TVLISGGTGLLGGLVARHLVAERGVRRLVLAGR----RGWSAPGVTDLVDELVGLGAAVE 1460

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAE 453
Db       1461 VA-----------SCDVGDRAQLDRLLTTISAE 1482


RESULT 4
US-09-237-119-2
; Sequence 2, Application US/09237119
  Patent No. 6448038
  GENERAL INFORMATION:
;  APPLICANT: Kallender, Howard
;  APPLICANT: Sylvester, Daniel R.
;  APPLICANT: McDevitt, Damien
;  APPLICANT: Huang, Jianzhong
;  APPLICANT: Van Horn, Stephanie
;  TITLE OF INVENTION: birA
;  FILE REFERENCE: GM10189
;  CURRENT APPLICATION NUMBER: US/09/237,119
;  CURRENT FILING DATE: 1999-01-26
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-237-119-2

  Query Match             5.2%;  Score 125;  DB 4;  Length 323;
  Best Local Similarity   27.8%;  Pred. No. 6.6e-05;
  Matches   44;  Conservative   25;  Mismatches   71;  Indels   18;  Gaps    6;

Qy         41 ETQIKWPNDLVVGRDKLGGILIETVRAGGKTVAVV-GIGINFVLPKEVEN--------AA 91
Db        172 EVKVKWPNDIYIDNGKVCGFLTEMVANNDGIEAIICGIGIN--LTQQLENFDESIRHRAT 229

Qy         92 SVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLL 151
Db        230 SIQ-----LHDKNKLDRYQFLERLLQEIEKRYNQFLTLPFSEIREEYIAASNIWNR-TLL 283

Qy        152 LRDGETVCEGTVKGVDGRGVLHLETAEGE-QTVVSGEI 188
Db        284 FTENDKQFKGQAIDLDYDGYLIVRDEAGESHRLISADI 321


RESULT 5
US-09-134-001C-3229
; Sequence 3229, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3229
;   LENGTH: 328
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3229

  Query Match             4.8%;  Score 115;  DB 4;  Length 328;
  Best Local Similarity   25.1%;  Pred. No. 0.00081;
  Matches   49;  Conservative   31;  Mismatches   73;  Indels   42;  Gaps    8;

Qy         16 PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGGILIETVRAGGKTVAV 74
Db        156 PKFNL-----FIALGIRDAIQQFSNERVTIKWPNDIYIGNKKICGFLTEMVANYDEIEAI 210

Qy         75 V-GIGINFVLPKEVENAASVQSLFQTASRRGNADAAV----LLETLLAELGAVLEQYAEE 129
Db        211 ICGIGINM---NHVESDFDEDIKDKATSIRMHSDSIINRYTFLTALLTQIIHRFDQFLHQ 267

Qy        130 GFAPFLNEY--------------ETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLE 175
Db        268 TFESIREEYIHATNIWHRQLKFTENNHQFLGEAIDIDSDGFLIVK------DEKGQLH-- 319

Qy        176 TAEGEQTVVSGEISL 190
Db        320 ------RLMSADIDL 328


RESULT 6
US-08-804-227C-5
; Sequence 5, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1611 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-5

  Query Match             4.6%;  Score 110;  DB 2;  Length 1611;
  Best Local Similarity   22.1%;  Pred. No. 0.04;
  Matches   85;  Conservative   42;  Mismatches  137;  Indels  120;  Gaps   19;

Qy         87 VENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYA-EEGFAPFLNEYETANRDH 145
Db        215 LHNAAQ-------ALRRGECDLALAGGVTVMATPAVFVQFARQRGLAP------------ 255
```

```
Qy        146 GKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGEQTVVSGEIS---------------- 189
                                                                          
Db        256 --------DGR--CKAFADAADGFG-----PAEGVGMVLVERLSDARRLGHPVLAVVCGS 300

Qy        190 -LRPDNRS--VSVPKRPDSERFLLLEGGNSRLKWAWVE----NGTFATVGSAPYRDLSPL 242
                                                                          
Db        301 AVNQDGASNGLTAPSGPSQERVIRQALGNARLTVADVDVVEAHGTGTRLGD-PIEAQALL 359

Qy        243 GAEWAEKADGNVRI--------VGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHY 294
                                                                          
Db        360 GT-YGRDRDGGRPVWLGSLKSNIGHAQAAAGVAGVIKMVLAMRYGWLPRTL-------HV 411

Qy        295 RHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLG------------G 342
                                                                          
Db        412 DEPSRHVD--W-------------------SAGGVRLLTEAREWPGVDRPRRAAVSAFG 449

Qy        343 TIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMM---HGRL 399
                                                                          
Db        450 VSGTNAHLILE--APDTAE-----AESATTPVRSEVSESAAVLDARSGVVPVVVSGRSRV 502

Qy        400 KEKNGAGKPVDVIITGG-GAAKVA 422
                                      
Db        503 VVREAAGRLAEVVEAGGVGLADVA 526
```

RESULT 7
US-08-095-734-2
; Sequence 2, Application US/08095734
; Patent No. 5807723
; GENERAL INFORMATION:
; APPLICANT: Aldovini, Anna
; APPLICANT: Young, Richard A.
; TITLE OF INVENTION: Homologously Recombinant Slow Growing
; TITLE OF INVENTION: Mycobacteria and Uses Therefor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/095,734
; FILING DATE: 22-JUL-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI93-11M
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1271 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-095-734-2

Query Match 4.3%; Score 101.5; DB 1; Length 1271;
Best Local Similarity 22.4%; Pred. No. 0.22;
Matches 101; Conservative 47; Mismatches 186; Indels 117; Gaps 20;

```
Qy         26 VAALACRRALGCLGLETQIK----WPNDLVVGRDKLGGILIETVRAGGKTVAVVGIGINF 81
                                                                          
Db        504 LAEAKARRGPLCLGIDPHPELLRGW--DLATTADGLAAFCDICVRAFA-DFAVVKPQVAF 560

Qy         82 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEE--GFAPFLNEYE 139
                                                                          
Db        561 F---ESYGAAGFAVLERTIAELRAADVLVLADAKRGDIGATMSAYATAWVGDSPLAADAV 617

Qy        140 TAN---------------RDHGKAVLLLR-----DGETVCEGTVKGVDGRGVLHLETAEG 179
                                                                          
Db        618 TASPYLGFGSLRPLLEVAAAHGRGVFVLAATSNPEGAAVQNA---AADGRSVAQLVV--- 671

Qy        180 EQTVVSGEISLRPDNRSVSV------PKRPDSERF---LLLEGGNSRLKWAWVENG---T 227
                                                                          
Db        672 DQVGAANE-AAGPGPGSIGVVVGATAPQAPDLSAFTGPVLVPG-------VGVQGGRPEA 723

Qy        228 FATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEW----LPS 283
                                                                          
Db        724 LGGLGGAASSQLLPAVAREVLRAGPGVPELRAA--GERMRDAVAYLAAVMWQWPLRGLRS 781

Qy        284 SAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 343
                                                                          
Db        782 SA---------------------ARGSGWPNRVPVMPLGCAG-------------GGV 805

Qy        344 IMPGFHLMKESLA----VRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRL 399
                                                                          
Db        806 VSQRWAMTEPWIAFPPEVHSAMLNYGAGVGPMLISATQNGELSAQYAEAASEVEELLG-- 863

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFL 430
                                           
Db        864 -----------VVASEGWQGQAAEALVAAYM 883
```

RESULT 8
US-08-444-623-2
; Sequence 2, Application US/08444623
; Patent No. 5866403
; GENERAL INFORMATION:
; APPLICANT: Aldovini, Anna
; APPLICANT: Young, Richard A.
; TITLE OF INVENTION: Homologously Recombinant Slow Growing
; TITLE OF INVENTION: Mycobacteria and Uses Therefor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,623
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/095,734
; FILING DATE: 22-JUL-1993
; APPLICATION NUMBER: 07/711,334
; FILING DATE: 06-JUN-1991
; APPLICATION NUMBER: 07/367,894
; FILING DATE: 19-JUN-1989
; APPLICATION NUMBER: PCT/US90/03451
; FILING DATE: 18-JUN-1990
; APPLICATION NUMBER: PCT/US89/02962
; FILING DATE: 07-JUL-1989
; APPLICATION NUMBER: 07/361,944
; FILING DATE: 05-JUN-1989
; APPLICATION NUMBER: 07/223,089
; FILING DATE: 22-JULY-1988
; APPLICATION NUMBER: 07/216,390

```
;       FILING DATE:  07-JUL-1988
;       APPLICATION NUMBER:  07/163,546
;       FILING DATE:  03-MAR-1988
;       APPLICATION NUMBER:  PCT/US88/00614
;       FILING DATE:  29-FEB-1988
;       APPLICATION NUMBER:  07/020,451
;       FILING DATE:  02-MAR-1987
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Granahan, Patricia
;       REGISTRATION NUMBER:  32,227
;       REFERENCE/DOCKET NUMBER:  WHI93-11MZ
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617-861-6240
;       TELEFAX:  617-861-9540
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1271 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-444-623-2

  Query Match             4.3%;  Score 101.5;  DB 2;  Length 1271;
  Best Local Similarity   22.4%;  Pred. No. 0.22;
  Matches  101;  Conservative   47;  Mismatches  186;  Indels  117;  Gaps   20;

Qy         26 VAALACRRALGCLGLETQIK----WPNDLVVGRDKLGGILIETVRAGGKTVAVVGIGINF 81
              :|    ||   |||::   :    |  ||    | |       |||     |||   : |
Db        504 LAEAKARRGPLCLGIDPHPELLRGW--DLATTADGLAAFCDICVRAFA-DFAVVKPQVAF 560

Qy         82 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEE--GFAPFLNEYE 139
                 |   ||    | :| :    ||  || :    ::|| :  ||    | :|   :
Db        561 F---ESYGAAGFAVLERTIAELRAADVLVLADAKRGDIGATMSAYATAWVGDSPLAADAV 617

Qy        140 TAN--------------RDHGKAVLLLR-----DGETVCEGTVKGVDGRGVLHLETAEG 179
              ||:                  ||: | :|      :|  |       ||| |  |
Db        618 TASPYLGFGSLRPLLEVAAAHGRGVFVLAATSNPEGAAVQNA---AADGRSVAQLVV--- 671

Qy        180 EQTVVSGEISLRPDNRSVSV------PKRPDSERF---LLLEGGNSRLKWAWVENG---T 227
              :|   : | :   |   |: |      |: ||   |   :|: |           |: |
Db        672 DQVGAANE-AAGPGPGSIGVVVGATAPQAPDLSAFTGPVLVPG-------VGVQGGRPEA 723

Qy        228 FATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEW----LPS 283
              :| |    | | |    :|   |  :  |  ||  :  |    |   :|    | |
Db        724 LGGLGGAASSQLLPAVAREVLRAGPGVPELRAA--GERMRDAVAYLAAVMWQWPLRGLRS 781

Qy        284 SAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 343
              ||                      | ||   :|   : : |                ||
Db        782 SA----------------------ARGSGWPNRVPVMPLGCAG-------------GGV 805

Qy        344 IMPGFHLMKESLA----VRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRL 399
              :   : : :|    | :| ||  ||      : | |   |         :  : |
Db        806 VSQRWAMTEPWIAFPPEVHSAMLNYGAGVGPMLISATQNGELSAQYAEAASEVEELLG-- 863

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFL 430
                      |: : |   : ||||  |::
Db        864 ----------VVASEGWQGQAAEALVAAYM 883


RESULT 9
US-08-471-869-2
; Sequence 2, Application US/08471869
; Patent No. 6022745
;  GENERAL INFORMATION:
;    APPLICANT:  Aldovini, Anna
;    APPLICANT:  Young, Richard A.
;    TITLE OF INVENTION:  Homologously Recombinant Slow Growing
;    TITLE OF INVENTION:  Mycobacteria and Uses Therefor
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,869
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08267
;      FILING DATE:  22-JUL-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/095,734
;      FILING DATE:  22-JUL-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/711,334
;      FILING DATE:  06-JUN-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/367,894
;      FILING DATE:  19-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US90/03451
;      FILING DATE:  18-JUN-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US89/02962
;      FILING DATE:  07-JUL-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/361,944
;      FILING DATE:  05-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/223,089
;      FILING DATE:  22-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/216,390
;      FILING DATE:  07-JUL-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/163,546
;      FILING DATE:  03-MAR-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US88/00614
;      FILING DATE:  29-FEB-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/020,451
;      FILING DATE:  02-MAR-1987
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI93-11MA2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1271 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-471-869-2

  Query Match             4.3%;  Score 101.5;  DB 3;  Length 1271;
  Best Local Similarity   22.4%;  Pred. No. 0.22;
  Matches  101;  Conservative   47;  Mismatches  186;  Indels  117;  Gaps   20;

Qy         26 VAALACRRALGCLGLETQIK----WPNDLVVGRDKLGGILIETVRAGGKTVAVVGIGINF 81
              :|    ||   |||::   :    |  ||    | |       |||     |||   : |
Db        504 LAEAKARRGPLCLGIDPHPELLRGW--DLATTADGLAAFCDICVRAFA-DFAVVKPQVAF 560
```

```
Qy         82 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEE--GFAPFLNEYE 139
Db        561 F---ESYGAAGFAVLERTIAELRAADVLVLADAKRGDIGATMSAYATAWVGDSPLAADAV 617

Qy        140 TAN---------------RDHGKAVLLLR-----DGETVCEGTVKGVDGRGVLHLETAEG 179
Db        618 TASPYLGFGSLRPLLEVAAAHGRGVFVLAATSNPEGAAVQNA---AADGRSVAQLVV--- 671

Qy        180 EQTVVSGEISLRPDNRSVSV------PKRPDSERF---LLLEGGNSRLKWAWVENG---T 227
Db        672 DQVGAANE-AAGPGPGSIGVVVGATAPQAPDLSAFTGPVLVPG-------VGVQGGRPEA 723

Qy        228 FATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEW----LPS 283
Db        724 LGGLGGAASSQLLPAVAREVLRAGPGVPELRAA--GERMRDAVAYLAAVMWQWPLRGLRS 781

Qy        284 SAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 343
Db        782 SA---------------------ARGSGWPNRVPVMPLGCAG-------------GGV 805

Qy        344 IMPGFHLMKESLA----VRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRL 399
Db        806 VSQRWAMTEPWIAFPPEVHSAMLNYGAGVGPMLISATQNGELSAQYAEAASEVEELLG-- 863

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFL 430
Db        864 -----------VVASEGWQGQAAEALVAAYM 883
```

RESULT 10
US-09-342-563-2
; Sequence 2, Application US/09342563
; Patent No. 6355486
; GENERAL INFORMATION:
; APPLICANT: Aldovini, Anna
; APPLICANT: Young, Richard A.
; TITLE OF INVENTION: Homologously Recombinant Slow Growing
; TITLE OF INVENTION: Mycobacteria and Uses Therefor
; FILE REFERENCE: WHI93-11MA3
; CURRENT APPLICATION NUMBER: US/09/342,563
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: US 08/471,869
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: US 08/095,734
; EARLIER FILING DATE: 1993-07-22
; EARLIER APPLICATION NUMBER: PCT/US94/08267
; EARLIER FILING DATE: 1994-07-22
; EARLIER APPLICATION NUMBER: US 07/711,334
; EARLIER FILING DATE: 1991-06-06
; EARLIER APPLICATION NUMBER: US 07/367,894
; EARLIER FILING DATE: 1989-06-19
; EARLIER APPLICATION NUMBER: PCT/US90/03451
; EARLIER FILING DATE: 1990-06-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1271
; TYPE: PRT
; ORGANISM: Mycobacterium bovis Bacille Calmette-Guerin
US-09-342-563-2

Query Match 4.3%; Score 101.5; DB 4; Length 1271;
Best Local Similarity 22.4%; Pred. No. 0.22;
Matches 101; Conservative 47; Mismatches 186; Indels 117; Gaps 20;

```
Qy         26 VAALACRRALGCLGLETQIK----WPNDLVVGRDKLGGILIETVRAGGKTVAVVGIGINF 81
Db        504 LAEAKARRGPLCLGIDPHPELLRGW--DLATTADGLAAFCDICVRAFA-DFAVVKPQVAF 560

Qy         82 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEE--GFAPFLNEYE 139
Db        561 F---ESYGAAGFAVLERTIAELRAADVLVLADAKRGDIGATMSAYATAWVGDSPLAADAV 617

Qy        140 TAN---------------RDHGKAVLLLR-----DGETVCEGTVKGVDGRGVLHLETAEG 179
Db        618 TASPYLGFGSLRPLLEVAAAHGRGVFVLAATSNPEGAAVQNA---AADGRSVAQLVV--- 671

Qy        180 EQTVVSGEISLRPDNRSVSV------PKRPDSERF---LLLEGGNSRLKWAWVENG---T 227
Db        672 DQVGAANE-AAGPGPGSIGVVVGATAPQAPDLSAFTGPVLVPG-------VGVQGGRPEA 723

Qy        228 FATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEW----LPS 283
Db        724 LGGLGGAASSQLLPAVAREVLRAGPGVPELRAA--GERMRDAVAYLAAVMWQWPLRGLRS 781

Qy        284 SAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 343
Db        782 SA---------------------ARGSGWPNRVPVMPLGCAG-------------GGV 805

Qy        344 IMPGFHLMKESLA----VRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRL 399
Db        806 VSQRWAMTEPWIAFPPEVHSAMLNYGAGVGPMLISATQNGELSAQYAEAASEVEELLG-- 863

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFL 430
Db        864 -----------VVASEGWQGQAAEALVAAYM 883
```

RESULT 11
PCT-US94-08267-2
; Sequence 2, Application PC/TUS9408267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Homologously Recombinant Slow Growing
; TITLE OF INVENTION: Mycobacteria and Uses Therefor
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08267
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/095,734
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI93-11MA PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1271 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-08267-2

Query Match 4.3%; Score 101.5; DB 5; Length 1271;
Best Local Similarity 22.4%; Pred. No. 0.22;
Matches 101; Conservative 47; Mismatches 186; Indels 117; Gaps 20;

```
Qy         26 VAALACRRALGCLGLETQIK----WPNDLVVGRDKLGGILIETVRAGGKTVAVVGIGINF 81
```

```
Db        504 LAEAKARRGPLCLGIDPHPELLRGW--DLATTADGLAAFCDICVRAFA-DFAVVKPQVAF 560

Qy         82 VLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAEE--GFAPFLNEYE 139

Db        561 F---ESYGAAGFAVLERTIAELRAADVLVLADAKRGDIGATMSAYATAWVGDSPLAADAV 617

Qy        140 TAN---------------RDHGKAVLLLR-----DGETVCEGTVKGVDGRGVLHLETAEG 179

Db        618 TASPYLGFGSLRPLLEVAAAHGRGVFVLAATSNPEGAAVQNA---AADGRSVAQLVV--- 671

Qy        180 EQTVVSGEISLRPDNRSVSV------PKRPDSERF---LLLEGGNSRLKWAWVENG---T 227

Db        672 DQVGAANE-AAGPGPGSIGVVVGATAPQAPDLSAFTGPVLVPG-------VGVQGGRPEA 723

Qy        228 FATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEW----LPS 283

Db        724 LGGLGGAASSQLLPAVAREVLRAGPGVPELRAA--GERMRDAVAYLAAVMWQWPLRGLRS 781

          284 SAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGT 343

Db        782 SA---------------------ARGSGWPNRVPVMPLGCAG-------------GGV 805

Qy        344 IMPGFHLMKESLA----VRTANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRL 399

Db        806 VSQRWAMTEPWIAFPPEVHSAMLNYGAGVGPMLISATQNGELSAQYAEAASEVEELLG-- 863

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFL 430

Db        864 -----------VVASEGWQGQAAEALVAAYM 883


RESULT 12
US-08-804-227C-2
; Sequence 2, Application US/08804227C
; Patent No. 5876991
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Kuhstoss, Stuart A.
;    APPLICANT:  Rosteck, Paul R., Jr.
;    APPLICANT:  Sutton, Kimberly L.
;    TITLE OF INVENTION:  POLYKETIDE SYNTHASE GENES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  THOMAS G. PLANT  1501
;      STREET:  LILLY CORPORATE CENTER
;      CITY:  INDIANAPOLIS
;      STATE:  IN
;      COUNTRY:  USA
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  MS-DOS
;      SOFTWARE:  ASCI(DOS) Text only
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/804,227C
;      FILING DATE:  February 21, 1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Plant, Thomas, G.
;      REGISTRATION NUMBER:  35,784
;      REFERENCE/DOCKET NUMBER:  X-8231
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-2459
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4472 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-804-227C-2
```

```
  Query Match             4.2%;  Score 100.5;  DB 2;  Length 4472;
  Best Local Similarity   22.6%;  Pred. No. 2.3;
  Matches   96;  Conservative   44;  Mismatches  163;  Indels  121;  Gaps   23;

Qy         69 GKTVAVVGIGINFVLPKE-----VENA--ASVQSLFQTAS--RRGNADAAVLLETLLAEL 119

Db       2721 GKSASVVSGRISYTLGLEGPSLSIDTACSSSLVALHNAAQALRGGECDMAL--------A 2772

Qy        120 GAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVD------GRGVLH 173

Db       2773 GGVTVMAAPGLFVEF-------SRQRG----LAADGR--CKAFADGADGTAWAEGAGVVL 2819

Qy        174 LETAEGEQ-------TVVSGEISLRPDNRS--VSVPKRPDSERFLLLEGGNSRLKWAWVE 224

Db       2820 VERLSDARRLGHPVLAVVCGS-AVNQDGASNGLTAPSGPSQERVIRQALANARLTVADVD 2878

Qy        225 ----NGTFATVGSAPYRDLSPLG-------AE---WAEKADGNVRIVGCAVCGESKKAQV 270

Db       2879 VVEAHGTGTRLGD-PIEAQALLGTYGRDRDAECPVWLGSLKSN---IGHAQAAAGVAGVI 2934

Qy        271 KEQLARKIEWLPSSAQALGIRNHYRHPEEH-----------------GSDR-------W 305

Db       2935 KMVLAMRYGWLPRTL-------HVDEPSRHVDWSAGGVRLLTEAREWPGVDRPRRAAVSA 2987

Qy        306 FNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRP 365

Db       2988 FGVSGT-----NAHLILEAPEALEALEATDAPEAPEAPEAPDVTDVTEALEA-------- 3034

Qy        366 AGKRYPFPTTTGNAVASGMMD------AVCGSIMMMHGRLKEKNGAGKPVDVIITGG-GA 418

Db       3035 -----PDATEAEGAKAPGSPEEAQPAVGVVPVVVSGRSRVVVREAAGRLAEVVEAGGVGL 3089

Qy        419 AKVA 422

Db       3090 ADVA 3093


RESULT 13
US-09-105-537-31
; Sequence 31, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 4613
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-31

  Query Match             4.2%;  Score 100;  DB 4;  Length 4613;
  Best Local Similarity   23.0%;  Pred. No. 2.7;
  Matches   98;  Conservative   60;  Mismatches  177;  Indels   92;  Gaps   21;

Qy         53 GRDKLGGILIETVRAGGKTVAVVGIGINFVLPKE-----VENAAS---------VQSLFQ 98

Db       1243 GGEGLDGYLLT-----GNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQAL-- 1295

Qy         99 TASRRGNADAAVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETV 158

Db       1296 ---RKGEVDMAL--------AGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSW 1344

Qy        159 CEGTVKGVDGRGVLHLE-----TAEGEQ--TVVSGEISLRPDNRS--VSVPKRPDSERFL 209

Db       1345 SEGV-------GVLLVERLSDARRNGHQVLAVVRGS-ALNQDGASNGLTAPNGPSQQRVI 1396
```

```
Qy        210 LLEGGNSRLKWAWVE----NGTFATVGS--------APY----RDLSPL----------G 243
              ::||  : |:    :||   :|          | |     |  ||
Db       1397 RRALADARLTTSDVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGH 1456

Qy        244 AEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSD 303
               : |    | :::|     |   |    :: : :|:|   : : |     :   :: |
Db       1457 TQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLR 1516

Qy        304 RWFNALGSRRFS-RNACVVVSCGTAVTVDALTDD----GHYLGGTIMPGFHLMKESLAVR 358
              |   |: |   |  || ||:     |   |   :    | :|| : | : :  :| |
Db       1517 R--AAVSSFGISGTNAHVVLEEAPVVVEGASVVEPSVGGSAVGGGVTP-WVVSAKSAAAL 1573

Qy        359 TANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGA 418
              | : | |   :     | :| | | :||   : ::  || : :: |      :  | ||
Db       1574 DAQIERLAA--FASRDRTDDADA-GAVDAGAVAHVLADGRAQFEHRA------VALGAGA 1624

Qy        419 AKVAEAL 425
               : :||
Db       1625 DDLVQAL 1631


RESULT 14
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
;  APPLICANT: Sherman, D.H.
;  APPLICANT: Liu, H.
;  APPLICANT: Xue, Y.
;  APPLICANT: Zhao, L.
;  TITLE OF INVENTION: DNA encoding methymycin and pikromycin
;  FILE REFERENCE: 600.438US1
;  CURRENT APPLICATION NUMBER: US/09/105,537A
;  CURRENT FILING DATE: 1998-06-26
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 6
;   LENGTH: 11877
;   TYPE: PRT
;   ORGANISM: Streptomyces venezuelae
US-09-105-537-6

  Query Match             4.2%;  Score 100;  DB 4;  Length 11877;
  Best Local Similarity   23.0%;  Pred. No. 13;
  Matches   98;  Conservative   60;  Mismatches  177;  Indels   92;  Gaps   21;

Qy         53 GRDKLGGILIETVRAGGKTVAVVGIGINFVLPKE-----VENAAS---------VQSLFQ 98
              | : |. | |:      | | :|:   ::: |  |     |: | |          ||:|
        1579 GGEGLDGYLLT-----GNTASVMSGRVSYTLGLEGPALTVDTACSSSLVALHLAVQAL-- 1631

Qy         99 TASRRGNADAAVLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETV 158
                |:|  | |:        | |      |  | :   |     ||    || :
Db       1632 ---RKGEVDMAL--------AGGVAVMPTPGMFVEFSRQRGLAGDGRSKAFAASADGTSW 1680

Qy        159 CEGTVKGVDGRGVLHLE-----TAEGEQ--TVVSGEISLRPDNRS--VSVPKRPDSERFL 209
              ||         ||| :|         | |   || |  :|  |  |  :: |  |  :| :
Db       1681 SEGV-------GVLLVERLSDARRNGHQVLAVVRGS-ALNQDGASNGLTAPNGPSQQRVI 1732

Qy        210 LLEGGNSRLKWAWVE----NGTFATVGS--------APY----RDLSPL----------G 243
               ::||  : |:    :||   :|          | |     |  ||
Db       1733 RRALADARLTTSDVDVVEAHGTGTRLGDPIEAQALIATYGQGRDDEQPLRLGSLKSNIGH 1792

Qy        244 AEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEHGSD 303
               : |    | :::|     |   |    :: : :|:|   : : |     :   :: |
Db       1793 TQAAAGVSGVIKMVQAMRHGLLPKTLHVDEPSDQIDWSAGAVELLTEAVDWPEKQDGGLR 1852

Qy        304 RWFNALGSRRFS-RNACVVVSCGTAVTVDALTDD----GHYLGGTIMPGFHLMKESLAVR 358
              |   |: |   |  || ||:     |   |   :    | :|| : | : :  :| |
Db       1853 R--AAVSSFGISGTNAHVVLEEAPVVVEGASVVEPSVGGSAVGGGVTP-WVVSAKSAAAL 1909

Qy        359 TANLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGA 418
              | : | |   :     | :| | | :||   : ::  || : :: |      :  | ||
Db       1910 DAQIERLAA--FASRDRTDDADA-GAVDAGAVAHVLADGRAQFEHRA------VALGAGA 1960

Qy        419 AKVAEAL 425
               : :||
Db       1961 DDLVQAL 1967


RESULT 15
US-09-605-785-826
; Sequence 826, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan L.
;  APPLICANT:  Jiang, Yuqui
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Stolk, John A.
;  APPLICANT:  Day, Craig H.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Li, Samuel
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Skeiky, Yasir A.W.
;  APPLICANT:  Hepler, William
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C16
;  CURRENT APPLICATION NUMBER: US/09/605,785
;  CURRENT FILING DATE:  2000-06-27
;  NUMBER OF SEQ ID NOS: 835
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 826
;   LENGTH: 357
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-605-785-826

  Query Match             4.1%;  Score 98.5;  DB 4;  Length 357;
  Best Local Similarity   24.2%;  Pred. No. 0.056;
  Matches   62;  Conservative   32;  Mismatches  105;  Indels   57;  Gaps    9;

Qy          9 FGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGILIETVRAG 68
              ||  ::: :||      : | | :: |       |:    || |  |:       | :|
Db        126 FGKKYEKERYE----KVIKACALKKDL-------QLLEDGDLTVIGDR------GTTLSG 168

Qy         69 GKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELGAVLEQYAE 128
              |:   |             | |  ::::|      :||  :| : | |    |     |
Db        169 GQKARV-------------NLA--RAVYQ------DADIYLLDDPLSAVDAEVSRHLFE 206

Qy        129 EGFAPFLNEYETANRDH-------GKAVLLLRDGETVCEGTV-----KGVDGRGVLHLET 176
                    |:|  |    |          :|:|:||: | :||       |:|   :|  :
Db        207 LCICQILHEKITILVTHQLQYLKAASQILILKDGKMVQKGTYTEFLKSGIDFGSLLKKDN 266

Qy        177 AEGEQTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWV------ENGTFAT 230
              | ||  | |  :||    | |      | |  | :|          |      || :
Db        267 EESEQPPVPGTPTLRNRTFSESSVWSQQSSRPSLKDGALESQDTENVPVTLSEENRSEGK 326

Qy        231 VGSAPYRDLSPLGAEW 246
              ||   |::    || |
Db        327 VGFQAYKNYFRAGAHW 342


Search completed: June 24, 2003, 21:49:40
Job time : 15.438 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 37.5177 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-39
Perfect score: 2384
Sequence: 1 MGECLMFSFGWAFDRPQYEL.....LVIHGLLNLIAAEGGESEHA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2223	93.2	592	16	Q9JWI7	Q9jwi7 neisseria m
2	2197	92.2	592	16	Q9JXF1	Q9jxf1 neisseria m
3	288	12.1	295	16	Q8Y2M4	Q8y2m4 ralstonia s
4	261.5	11.0	291	16	Q8Y2M3	Q8y2m3 ralstonia s
5	233.5	9.8	312	16	Q9HWC0	Q9hwc0 pseudomonas
6	232.5	9.8	248	16	Q9HWC1	Q9hwc1 pseudomonas
7	218.5	9.2	288	2	Q9ZIU6	Q9ziu6 bordetella
8	214.5	9.0	320	16	Q9KV39	Q9kv39 vibrio chol
9	210	8.8	321	16	Q8X709	Q8x709 escherichia
10	199.5	8.4	242	16	Q9PCI4	Q9pci4 xylella fas
11	195	8.2	312	16	Q9CNX6	Q9cnx6 pasteurella
12	194	8.1	319	16	Q8ZAN5	Q8zan5 yersinia pe
13	190	8.0	325	16	Q92AA4	Q92aa4 listeria in
14	187	7.8	320	16	Q8Z317	Q8z317 salmonella
15	187	7.8	325	16	Q8Y5Z9	Q8y5z9 listeria mo
16	185	7.8	261	16	Q9A6Z1	Q9a6z1 caulobacter
17	185	7.8	273	16	Q97EB4	Q97eb4 clostridium
18	179	7.5	265	16	Q9X8N6	Q9x8n6 streptomyce
19	170.5	7.2	235	17	O73952	O73952 pyrococcus
20	170.5	7.2	317	16	Q9PCI3	Q9pci3 xylella fas
21	166	7.0	255	16	Q92QN8	Q92qn8 rhizobium m
22	166	7.0	257	2	Q9R9L1	Q9r9l1 rhizobium m
23	164	6.9	255	16	Q8R7M2	Q8r7m2 thermoanaer
24	162	6.8	274	16	P72816	P72816 synechocyst
25	162	6.8	326	16	Q8R8N0	Q8r8n0 thermoanaer
26	161	6.8	322	2	O32511	O32511 deinococcus
27	161	6.8	474	16	Q9RRZ1	Q9rrz1 deinococcus
28	160.5	6.7	233	16	O66837	O66837 aquifex aeo
29	157.5	6.6	259	16	Q8XHL5	Q8xhl5 clostridium
30	157	6.6	274	16	Q9CCL3	Q9ccl3 mycobacteri
31	153.5	6.4	274	16	Q9CD56	Q9cd56 mycobacteri
32	153	6.4	258	2	Q9F985	Q9f985 bacillus st
33	152.5	6.4	246	16	Q9WZY5	Q9wzy5 thermotoga
34	152	6.4	282	17	Q9HRK0	Q9hrk0 halobacteri
35	151.5	6.4	237	17	Q9V2C7	Q9v2c7 pyrococcus
36	151.5	6.4	242	2	Q52660	Q52660 rhodobacter
37	149	6.2	256	16	Q8UFW3	Q8ufw3 agrobacteri
38	148.5	6.2	250	16	Q9CED9	Q9ced9 lactococcus
39	147.5	6.2	311	16	Q97NV9	Q97nv9 streptococc
40	147.5	6.2	333	1	Q977S0	Q977s0 uncultured
41	146.5	6.1	272	16	O06282	O06282 mycobacteri
42	146.5	6.1	273	16	O83446	O83446 treponema p
43	145.5	6.1	323	16	Q9KC88	Q9kc88 bacillus ha
44	143.5	6.0	229	1	Q977N1	Q977n1 uncultured
45	143	6.0	261	17	O27938	O27938 methanobact

ALIGNMENTS

RESULT 1
Q9JWI7
ID Q9JWI7 PRELIMINARY; PRT; 592 AA.
AC Q9JWI7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BirA bifunctional protein (EC 6.3.4.15).
GN BIRA OR NMA0357.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z2491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162752; CAB83659.1; -.
DR HSSP; P06709; 1BIA.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BirA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00121; birA_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Ligase; Complete proteome.
SQ SEQUENCE 592 AA; 64468 MW; 7ED2E0CD3B31C630 CRC64;

```
  Query Match             93.2%;  Score 2223;  DB 16;  Length 592;
  Best Local Similarity   93.7%;  Pred. No. 2.6e-147;
  Matches  430;  Conservative   12;  Mismatches   17;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||| |||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 2
Q9JXF1
ID   Q9JXF1      PRELIMINARY;      PRT;   592 AA.
AC   Q9JXF1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein/Bvg accessory factor.
GN   NMB2075.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
     NCBI_TaxID=491;
     [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002557; AAF42394.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; NMB2075; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   592 AA;  64701 MW;  9E27421DA2B41DE3 CRC64;

  Query Match             92.2%;  Score 2197;  DB 16;  Length 592;
  Best Local Similarity   92.6%;  Pred. No. 1.7e-145;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||: ||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||| |||||||||:||||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| |||||||:|||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||||||||||||| |||||:|||||||||||||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||:||||:||||| | ||
Db        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 3
Q8Y2M4
ID   Q8Y2M4      PRELIMINARY;      PRT;   295 AA.
AC   Q8Y2M4;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcription regulation accessory factor transcription
DE   regulator protein.
GN   RSC0311 OR RS03278.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13839.1; -.
DR   InterPro; IPR004619; Baf.
```

```
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Complete proteome.
SQ   SEQUENCE   295 AA;  30990 MW;  133074A7764BFCBB CRC64;

  Query Match           12.1%;  Score 288;  DB 16;  Length 295;
  Best Local Similarity   30.7%;  Pred. No. 2.4e-12;
  Matches   90;  Conservative   50;  Mismatches  105;  Indels   48;  Gaps   12;

Qy        201 KRPDSER-FLLLEGGNSRLKWAW----------VENGT-FATVGSAPYRDLSPLGAEW-- 246
              | |  ||  ||:: ||:|:|||||             || :   |: |:  |: |  :|
Db          5 KAPTPERPLLLIDAGNTRIKWAWTAADVAPPAVAPGGTPWQHAGARPHDQLAELVEDWRD 64

Qy        247 ----AEKADGNVRIVGCAVCGESKKAQVKEQLAR-----KIEWLPSSAQALGIRNHYRHP 297
                  |  :| |    | | : :  :  ::||     ::  : | | | |:|| || |
Db         65 CHAGAGMAPPDVWI--SVVAGPALRDALCARIARVFDGARLRIVASEAAAAGLRNGYRDP 122

Qy        298 EEHGSDRWFNALGSRR-FSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLA 356
               : |:|||  |:|:|  :   | ::|: ||| |:| :  || : || |:|| ||  :|:
Db        123 AQLGTDRWVGAVGARHAWPDTALLLVTAGTATTLDIVAPDGRFAGGLILPGLTLMMRALS 182

Qy        357 VRTANL-----------NRPAGKRYP-FPTTTGNAVASGMMDAVCGSIMMMHGRLKEKN 403
                | |           :  |    | :   | :|:| | : |  |:|      |: :
Db        183 RNTAQLPEIDIGYLAARDDAQAPADVPSWADNTQDAIALGCVTAQAGAIAQTWQALQAQY 242

Qy        404 GAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGE 456
                |   :::||  | :|  |   |      :: ||||: | | ::|  | |
Db        243 PG--PYRCVLSGGARAALAPHLRMPF------QMHDNLVLLG-LQVLAHTGHE 286


RESULT 4
Q8Y2M3
ID   Q8Y2M3      PRELIMINARY;      PRT;   291 AA.
AC   Q8Y2M3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative bifunctional protein: biotin operon repressor and
DE   biotin--[acetyl-CoA-carboxylase] synthetase (EC 6.3.4.15).
GN   BIRA OR RSC0312 OR RS03279.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL   Nature 415:497-502(2002).
DR   EMBL; AL646058; CAD13840.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   291 AA;  30057 MW;  B9F1FD254D8F18C7 CRC64;

  Query Match           11.0%;  Score 261.5;  DB 16;  Length 291;
  Best Local Similarity   35.9%;  Pred. No. 1.7e-10;
  Matches   70;  Conservative   30;  Mismatches   86;  Indels    9;  Gaps    4;

Qy          5 LMFSFGWAFDRPQYELGSLSPVAALACRRALGCL----GLETQ---IKWPNDLVVGRDKL 57
              | ||   ||     :|  ||    :|   |:       |   |   :||||||: :   ||
Db         91 LTFSVACAFAAEPAQLAGLSLAVGVAVAEAVAAYAQTHGGNAQALALKWPNDVQIAGRKL 150

Qy         58 GGILIETVRAG-GKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLL 116
              |||:||:||  |:| ||:|||:|   |  :| ||  :||           :  :|  ||
Db        151 AGILVETMRAAPGQTWAVIGIGLNLERPHALE-AALGRSLSGVEELVERPEPNAVLAALL 209

Qy        117 AELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLET 176
              : ||  |:::   | |||:  :   :   |: | | :||  | ||   |:| :| | :|:
Db        210 SALGEHLQRFGSHGLAPFVAPFAARDAFAGRPVRLWQDGAVVLEGVAHGIDAQGRLAIES 269

Qy        177 AEGEQTVVSGEISLR 191
                | : |||:|||
Db        270 GGRVQWIHSGEVSLR 284


RESULT 5
Q9HWC0
ID   Q9HWC0      PRELIMINARY;      PRT;   312 AA.
AC   Q9HWC0;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein.
GN   BIRA OR PA4280.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
DR   EMBL; AE004843; AAG07668.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   312 AA;  34049 MW;  AC92F3BA43C4BCAF CRC64;

  Query Match             9.8%;  Score 233.5;  DB 16;  Length 312;
  Best Local Similarity   36.9%;  Pred. No. 1.7e-08;
  Matches   72;  Conservative   23;  Mismatches   89;  Indels   11;  Gaps    5;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGL-ETQIKWPNDLVVGRDKLGGI 60
              || | :|           ||  || |  ||  |||  ||: :| :|||||: |  |: ||
Db        121 GENLYYSLAITVRGGARELEGLSLVVGLAVLRALESLGVQDTGLKWPNDLLQGGKKIAGI 180

Qy         61 LIETVRAGGKTV----AVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLL 116
              |:|     |        |:|||||  :    :      || :        | : ||  |
Db        181 LLE---LSGDPADLCHVVIGIGINVNMKTATDIGQPWTSLREALGEL--VDRSRLLSALN 235

Qy        117 AELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLET 176
               :|   |:::: ||||    |:|: |   |  |:|     :: || | ||| || | ||
Db        236 RQLADYLDRHSREGFAASREEWESCNLWQGAQVILSTALSSI-EGRVLGVDSRGALRLEV 294

Qy        177 AEGEQTVVSGEISLR 191
                ||    ||:|||
Db        295 DGVEQGFSGGELSLR 309


RESULT 6
Q9HWC1
ID   Q9HWC1      PRELIMINARY;      PRT;   248 AA.
AC   Q9HWC1;
```

```
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein PA4279.
GN   PA4279.
OS   Pseudomonas aeruginosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=287;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 15692 / PAO1;
RX   MEDLINE=20437337; PubMed=10984043;
RA   Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA   Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA   Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA   Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA   Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA   Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT   "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT   opportunistic pathogen.";
RL   Nature 406:959-964(2000).
RR   EMBL; AE004843; AAG07667.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR001230; Prenyl_site.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   248 AA;  26758 MW;  609D37480899DA3B CRC64;

  Query Match             9.8%;  Score 232.5;  DB 16;  Length 248;
  Best Local Similarity   27.2%;  Pred. No. 1.4e-08;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy        209 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 268
              | |: ||| :||  :|    :  |     | : :    :::|   : :  | :     :
Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Qy        269 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 321
              :  : :||  :  | ||       |  |:|| |   :  | |||   : :   :: ||:|
Db         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Qy        322 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 374
              :  ||||| | :  || :||| | ||  ||:  |   |       |   |       |
Db        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Qy        375 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 429
                | |  :     | ::|: | ::|:        |   :: :||| |  | : |  |
Db        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

          430 LAENTVRVADNLVIHGL 446
                    |:  :||  ||
Db        231 ------RIMPDLVFVGL 241


RESULT 7
Q9ZIU6
ID   Q9ZIU6      PRELIMINARY;      PRT;   288 AA.
AC   Q9ZIU6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative biotin protein ligase.
GN   BIRA.
OS   Bordetella pertussis.
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=520;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP504;
RA   Wood G.E., Friedman R.L.;
RT   "Identification of a birA homolog in Bordetella pertussis.";
RL   Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF016461; AAC68833.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Ligase.
SQ   SEQUENCE   288 AA;  29666 MW;  39F6E9CDBC5F6F77 CRC64;

  Query Match             9.2%;  Score 218.5;  DB 2;  Length 288;
  Best Local Similarity   31.5%;  Pred. No. 1.7e-07;
  Matches   64;  Conservative   35;  Mismatches   91;  Indels   13;  Gaps    4;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCL-----GLETQIKWPNDLVVGRD 55
              :|  ||||  :  | |  :| :|||:| :|   ||  |     ||   :|||||:
Db         86 VGAALMFSCAFDVDLPPSQLPALSPLAGMAACEALRQLVGQPAGL--LVKWPNDVQWHDA 143

Qy         56 KLGGILIETVR--AGGKTVAVVGIGINF----VLPKEVENAASVQSLFQTASRRGNADAA 109
              || |:|:|:||  ||     |:|:|:|      | : :: : :  :       :|   ||
Db        144 KLAGVLVESVRNPAGTGHAVVIGMGLNLHGGDALSQALQRSVADWNDVAGPGEKGPDAAA 203

Qy        110 VLLETLLAELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGR 169
               |:  :             :|||  |   :   :   |: | ::  || :  |   |||
Db        204 DLVAAVARAWQDAAHTLQQEGFGAFGARFARVDALAGRPVNVIDRGEILYTGAASGVDDH 263

Qy        170 GVLHLETAEGEQTVVSGEISLRP 192
              | | :: |:| : |  ||||:||
Db        264 GRLLVQAADGMRPVSVGEISIRP 286


RESULT 8
Q9KV39
ID   Q9KV39      PRELIMINARY;      PRT;   320 AA.
AC   Q9KV39;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein.
GN   VC0319.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004120; AAF93492.1; -.
DR   HSSP; P06709; 1BIA.
DR   TIGR; VC0319; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   320 AA;  35040 MW;  88E6C496EBEFB5B6 CRC64;

  Query Match             9.0%;  Score 214.5;  DB 16;  Length 320;
  Best Local Similarity   31.8%;  Pred. No. 3.7e-07;
```

```
  Matches   62;  Conservative   27;  Mismatches  103;  Indels    3;  Gaps    3;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGGI 60
              |  |  |  |  |         || |  :|   ||  :|::  ::||||||     || ||
Db        127 GANLYLSIYWRLDAGMAAAMGLSLVVGVAIVEALEAMGVDGVKLKWPNDLYYQDKKLAGI 186

Qy         61 LIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 119
              |:| : :||     |:|:|||  :    |          : :   |    |    |:| |
Db        187 LVEMSGQAGAAAHLVIGMGINLAMRDNEGNIDQPWISLAEVTGQSRIDRNALAINLIAAL 246

Qy        120 GAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEG 179
                | ||   |   |:  :   :   |: | ||  |     |  :|:|  | : |||  ||
Db        247 DRTLRQYEISGMQNFVERWNRWDNFIGRPVKLLM-GANEVRGIERGIDEHGGVLLETEEG 305

Qy        180 EQTVVSGEISLRPDN 194
              :: : ||||||  ::
Db        306 LKSFIGGEISLRKND 320


 ESULT 9
Q8X709
ID   Q8X709      PRELIMINARY;      PRT;   321 AA.
AC   Q8X709;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Biotin-[acetylCoA carboxylase] holoenzyme synthetase and biotin operon
DE   repressor (Bifunctional protein BirA biotin-[acetylCoA carboxylase]
DE   holoenzyme synthetase / biotin operon repressor).
GN   BIRA OR Z5544 OR ECS4900.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
DR   EMBL; AE005629; AAG59173.1; -.
DR   EMBL; AP002567; BAB38323.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   321 AA;  35329 MW;  E1C1BDDD39E2B68D CRC64;

  Query Match             8.8%;  Score 210;  DB 16;  Length 321;
  Best Local Similarity   33.7%;  Pred. No. 7.7e-07;
  Matches   65;  Conservative   29;  Mismatches   93;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| |||  : : || :   |          | |   |   |: |
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAGINLDRNTLAAMLIRE 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| ||:|: :|  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLAPYLSRWEKLDNFINRPVKLI-IGDKEIFGISRGIDKQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 IIKPWMGGEISLR 317


RESULT 10
Q9PCI4
ID   Q9PCI4      PRELIMINARY;      PRT;   242 AA.
AC   Q9PCI4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Xf1795.
GN   XF1795.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004001; AAF84603.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   242 AA;  25658 MW;  EDC5DE7C0449C1E9 CRC64;

  Query Match             8.4%;  Score 199.5;  DB 16;  Length 242;
  Best Local Similarity   30.3%;  Pred. No. 2.8e-06;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;
```

```
Qy        210 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 269
              | : |||| | | :  |    |   ||   :        |   | | :  :|   :
Db          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH 63

Qy        270 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 326
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121

Qy        327 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 386
              |:|:| |  :| :|||  |     ||:::|  |   |    |    |   | :|: ||
Db        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181

Qy        387 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 442
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228

Qy        443 IHGL 446
              : ||
Db        229 LDGL 232


ESULT 11
Q9CNX6
ID   Q9CNX6      PRELIMINARY;      PRT;   312 AA.
AC   Q9CNX6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA.
GN   BIRA OR PM0296.
OS   Pasteurella multocida.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Pasteurella.
OX   NCBI_TaxID=747;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PM70;
RX   MEDLINE=21145866; PubMed=11248100;
RA   May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT   "Complete genomic sequence of Pasteurella multocida Pm70.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR   EMBL; AE006065; AAK02380.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   312 AA;  35214 MW;  0DD8B737C7C0E29B CRC64;

  Query Match             8.2%;  Score 195;  DB 16;  Length 312;
  Best Local Similarity   31.5%;  Pred. No. 8.3e-06;
  Matches   62;  Conservative   33;  Mismatches   86;  Indels   16;  Gaps    8;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLG-LETQIKWPNDLVVGRDKLGGI 60
              |: :| |  |  |: :  |  || |  :|    |   | |   :|||||:::   || ||
Db        122 GQVIM-SLYWTCDQ-KVNLEGLSLVVGMAIAETLKQAGALNIGLKWPNDVLLHGRKLAGI 179

Qy         61 LIETV-RAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 119
              |:|   |   :   |:| |||   ||:       | ::|  :        :    |:|||
Db        180 LVEIANRQNNQHNLVIGFGINLSFPKQ------TQQIYQPWAELIEVLPTIDRNKLIAEL 233

Qy        120 G----AVLEQYAEEGF-APFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHL 174
              |    | |: : : |    |  ::   :   |: | :| : : :  |  :|:| || | |
Db        234 GKKLIARLQHFEQAGIDEAFRQDWRDMDAYFGEEVHVLTENQKI-TGIAQGIDKRGYLQL 292

Qy        175 ETAEGEQTVVSGEISLR 191
              |:|| |    ||:|||
Db        293 ITSEGIQYFNGGEVSLR 309


RESULT 12
Q8ZAN5
ID   Q8ZAN5      PRELIMINARY;      PRT;   319 AA.
AC   Q8ZAN5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional protein BirA (EC 6.3.4.15).
GN   BIRA OR YPO3759.
OS   Yersinia pestis.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Yersinia.
OX   NCBI_TaxID=632;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX   MEDLINE=21470413; PubMed=11586360;
RA   Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA   Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA   Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA   Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA   Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA   Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA   Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT   "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL   Nature 413:523-527(2001).
DR   EMBL; AJ414158; CAC93227.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   319 AA;  35164 MW;  B58E9306939E3B2A CRC64;

  Query Match             8.1%;  Score 194;  DB 16;  Length 319;
  Best Local Similarity   29.7%;  Pred. No. 1e-05;
  Matches   57;  Conservative   30;  Mismatches  101;  Indels    4;  Gaps    4;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGGI 60
              |  |  |  |  ::        || |  :     |  || |  ::|||||| :   || ||
Db        128 GANLYLSMYWRLEQGPAAAMGLSLVVGIVMAEVLQNLGAEKVRVKWPNDLYLNDKKLAGI 187

Qy         61 LIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 119
              |:| | : |     |:| |||  :     || : |            |   |   :|:||
Db        188 LVELTGKTGDAAQLVIGAGINLTMRGSETNAIN-QDWINLQDAGVTIDRNKLTAEILSEL 246

Qy        120 GAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEG 179
                : ::  || : ||: ::  :    : | |:   : :  |  :|:| :| | ||
Db        247 RLAVVKFENEGLSAFLSRWQKMDNYLNRPVKLIIGNQEIF-GIARGIDQQGALLLEQDGN 305

Qy        180 EQTVVSGEISLR 191
              :  : ||||||
Db        306 IKPYIGGEISLR 317


RESULT 13
Q92AA4
ID   Q92AA4      PRELIMINARY;      PRT;   325 AA.
AC   Q92AA4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein.
GN   BIRA OR LIN2018.
OS   Listeria innocua.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1642;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=CLIP 11262 / SEROVAR 6A;
RX   PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL596170; CAC97248.1; -.
DR   ListiList; LIN02018; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   325 AA;  36218 MW;  650536C597F781D9 CRC64;

  Query Match             8.0%;  Score 190;  DB 16;  Length 325;
  Best Local Similarity   31.3%;  Pred. No. 2e-05;
  Matches   62;  Conservative   37;  Mismatches   55;  Indels   44;  Gaps   12;

Qy         16 PQYELGSLSPVAALACRRAL-GCLGLETQIKWPNDLVVGRDKLGGILIETVRAGGKTV-- 72
              ||:       :|:||   |:     || :||||||: :|: |: |:| | ::|  :|:
Db        151 PQFTF-----IASLAITEAIENITKLEPKIKWPNDIYIGKRKICGVLTE-MQAEAETIHA 204

Qy         73 AVVGIGINF---VLPKEV-ENAASVQ-SLFQTASRRGNADAAVLLETLLAELGAVLEQYA 127
              ::|:|||      |:|: | |:|::  | :: ||:       ||: :|  |    | :
Db        205 VIIGMGINVNQQEFPEEIKEKASSLKLELGESVSRK------ALLQAILTSLEKYYELFL 258

Qy        128 EEGFAPF----------LNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETA 177
              ::||||            |  ||:  ||            | |||:   ||| |: :
Db        259 DKGFAPIKLLWETKAIPFGEKLTASTTKGK----------IHGQVKGISDEGVLLLQDS 307

Qy        178 EGE-QTVVSGEISLRPDN 194
               ||  :: | :| |  ||
Db        308 LGEIHSIYSADILL--DN 323


RESULT 14
Q8Z317
ID   Q8Z317      PRELIMINARY;      PRT;   320 AA.
AC   Q8Z317;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Bifunctional protein: biotin operon repressor and biotin-[acetyl-CoA
DE   carboxylase] synthetase.
GN   STY3741.
OS   Salmonella typhi.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=601;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CT18;
RX   MEDLINE=21534947; PubMed=11677608;
RA   Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA   Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA   Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA   Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA   Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA   Whitehead S., Barrell B.G.;
RT   "Complete genome sequence of a multiple drug resistant Salmonella
RT   enterica serovar Typhi CT18.";
RL   Nature 413:848-852(2001).
DR   EMBL; AL627279; CAD09497.1; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR004409; HTH_repressor.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   TIGRFAMs; TIGR00122; birA_repr_reg; 1.
KW   Complete proteome.
SQ   SEQUENCE   320 AA;  35396 MW;  B584A48BABBEF59F CRC64;

  Query Match             7.8%;  Score 187;  DB 16;  Length 320;
  Best Local Similarity   31.6%;  Pred. No. 3.1e-05;
  Matches   61;  Conservative   32;  Mismatches   94;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRELGADKVRVKWPNDLYLLDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| |||  : : ||     |          |  :|   |: :
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGITLDRNMLAAKLIYK 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| :|:|: ::  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLI-IGDKEIFGISRGIDTQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
              :  : ||||||
Db        305 VIKPWMGGEISLR 317


RESULT 15
Q8Y5Z9
ID   Q8Y5Z9      PRELIMINARY;      PRT;   325 AA.
AC   Q8Y5Z9;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA protein.
GN   BIRA OR LMO1904.
OS   Listeria monocytogenes.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC   Listeriaceae; Listeria.
OX   NCBI_TaxID=1639;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EGD-E / SEROVAR 1/2A;
RX   MEDLINE=21537279; PubMed=11679669;
RA   Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA   Baquero F., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA   Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA   Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA   Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA   Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA   Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA   Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA   Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA   Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA   Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT   "Comparative genomics of Listeria species.";
RL   Science 294:849-852(2001).
DR   EMBL; AL591981; CAC99982.1; -.
DR   ListiList; LMO01904; -.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   Pfam; PF02237; BPL_C; 1.
```

```
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
KW   Complete proteome.
SQ   SEQUENCE   325 AA;  36236 MW;  EC954040194CDB21 CRC64;

  Query Match             7.8%;  Score 187;  DB 16;  Length 325;
  Best Local Similarity   30.8%;  Pred. No. 3.2e-05;
  Matches   61;  Conservative   35;  Mismatches   58;  Indels   44;  Gaps   11;

Qy         16 PQYELGSLSPVAALACRRAL-GCLGLETQIKWPNDLVVGRDKLGGILIETVRAGGKTV-- 72
              ||:       :|:||   |:     || :||||||: :|: |: |:| | ::|  :|:
Db        151 PQFTF-----IASLAITEAIENITKLEPKIKWPNDIYIGKRKICGVLTE-MQAEAETIHA 204

Qy         73 AVVGIGINF---VLPKEVENAAS--VQSLFQTASRRGNADAAVLLETLLAELGAVLEQYA 127
              ::|:|||      |:|::: ||     | :: ||:       ||: :|  |    | :
Db        205 VIIGMGINVNQQEFPEEIKDKASSLKLELGESVSRK------ALLQEILTSLEKYYELFL 258

Qy        128 EEGFAPF----------LNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETA 177
              ::||||             |  ||:   ||             | |||:   ||| |:
Db        259 DKGFAPIKLLWETKAIPFGEKLTASTTKGK----------IHGQVKGISDEGVLLLQDT 307

Qy        178 EGE-QTVVSGEISLRPDN 194
              ||  :: | :| |  ||
Db        308 LGEVHSIYSADILL--DN 323


Search completed: June 24, 2003, 21:59:40
Job time : 38.5177 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:36:04 ; Search time 16.7781 Seconds
(without alignments)
2635.685 Million cell updates/sec

Title: US-09-813-453A-39
Perfect score: 2384
Sequence: 1 MGECLMFSFGWAFDRPQYEL.....LVIHGLLNLIAAEGGESEHA 460

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2223	93.2	592	2	H82031	probable biotin-[a
2	2197	92.2	592	2	B81009	BirA protein/Bvg a
3	258.5	10.8	267	2	I40327	baf protein - Bord
4	233.5	9.8	312	2	A83112	BirA bifunctional
5	232.5	9.8	248	2	H83111	hypothetical prote
6	214.5	9.0	320	2	E82337	birA bifunctional
7	210	8.8	321	1	BVECBF	biotin-[acetyl-CoA
8	210	8.8	321	2	D91241	hypothetical prote
9	210	8.8	321	2	A86089	hypothetical prote
10	199.5	8.4	242	2	A82637	conserved hypothet
11	194	8.1	319	2	AG0457	biotin-[acetyl-CoA
12	190	8.0	325	2	AH1684	transcription regu
13	187	7.8	320	2	AG0934	bifunctional prote
14	187	7.8	325	2	AH1312	transcription regu
15	185	7.8	261	2	B87489	transcription acti
16	185	7.8	273	2	E97293	probable transcrip
17	179	7.5	265	2	T36391	hypothetical prote
18	170.5	7.2	317	2	B82637	bifunctional bioti
19	167	7.0	240	2	B47751	biotin [acetyl-CoA
20	162	6.8	274	2	S74680	biotin-[acetyl-CoA
21	161	6.8	325	2	A69595	transcription repr
22	161	6.8	474	2	D75285	birA bifunctional
23	160.5	6.7	233	2	A70351	biotin [acetyl-CoA
24	157	6.6	274	2	E87000	biotin apo-protein
25	153.5	6.4	274	2	H86937	conserved hypothet
26	152.5	6.4	246	2	D72320	conserved hypothet
27	152	6.4	282	2	B84224	biotin acetyl-CoA
28	151.5	6.4	237	2	H75202	biotin operon repr
29	149	6.2	256	2	AD2734	birA bifunctional
30	149	6.2	256	2	F97515	birA protein (AF20
31	148.5	6.2	250	2	G86862	biotin-[acetyl-CoA
32	147.5	6.2	311	2	B95222	BirA bifunctional
33	147.5	6.2	311	2	A99086	bioti-[acetyl-CoA-
34	146.5	6.1	272	2	A70955	hypothetical prote
35	146.5	6.1	273	2	D71326	conserved hypothet
36	145.5	6.1	323	2	E83860	biotin operon tran
37	143	6.0	261	2	H69122	biotin acetyl-CoA
38	139.5	5.9	274	2	AH2462	biotin acetyl-CoA
39	137.5	5.8	235	2	A71236	probable biotin-[a
40	137	5.7	254	2	F83660	hypothetical prote
41	137	5.7	257	2	S75559	hypothetical prote
42	135	5.7	250	2	C87489	biotin protein lig
43	134	5.6	233	2	S66100	conserved hypothet
44	134	5.6	323	2	D86847	biotin-[acetyl-CoA
45	132.5	5.6	1655	2	E97835	hypothetical prote

ALIGNMENTS

RESULT 1
H82031
probable biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Neisseria meningitid
N;Contains: biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15)
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 03-Jun-2002
C;Accession: H82031
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; M
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandr
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria menigitidis Z249
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: H82031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <PAR>
A;Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83659.1; PID:g73
A;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: birA; NMA0357
C;Keywords: ligase

Query Match 93.2%; Score 2223; DB 2; Length 592;
Best Local Similarity 93.7%; Pred. No. 2.7e-160;
Matches 430; Conservative 12; Mismatches 17; Indels 0; Gaps 0;

```
Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :|||||||||| |||||||||||||||:|||||  |||:||||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLKTQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| ||||||||:|||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

Qy        241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              ||||||||| |||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        373 PLGAEWAEKVDGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||:|||||||||| ||||||||||||||||||
```

```
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              ||||||||||||||||||||||||||||||||||||||
Db        553 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 591


RESULT 2
B81009
BirA protein/Bvg accessory factor NMB2075 [imported] - Neisseria meningitidis (strain MC
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: B81009
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: B81009
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-592 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42394.1; PID:g722733
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2075

  Query Match             92.2%;  Score 2197;  DB 2;  Length 592;
  Best Local Similarity   92.6%;  Pred. No. 2.5e-158;
  Matches  425;  Conservative   14;  Mismatches   20;  Indels    0;  Gaps    0;

Qy          1 MGECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLETQIKWPNDLVVGRDKLGGI 60
              :||||||||||| ||||||||||||||||:|||||  |||: |||||||||||||||||
Db        133 LGECLMFSFGWVFDRPQYELGSLSPVAAVACRRALSRLGLDVQIKWPNDLVVGRDKLGGI 192

Qy         61 LIETVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAELG 120
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        193 LIETVRTGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLVELD 252

Qy        121 AVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEGE 180
              ||| ||| :|||||: ||: |||||||||||||||||| |||||||||:||||||||:
Db        253 AVLLQYARDGFAPFVAEYQAANRDHGKAVLLLRDGETVFEGTVKGVDGQGVLHLETAEGK 312

Qy        181 QTVVSGEISLRPDNRSVSVPKRPDSERFLLLEGGNSRLKWAWVENGTFATVGSAPYRDLS 240
              |||||||||| |:| |||||| ||||||||:||||||||||||||||||||||||||||
Db        313 QTVVSGEISLRSDDRPVSVPKRRDSERFLLLDGGNSRLKWAWVENGTFATVGSAPYRDLS 372

          241 PLGAEWAEKADGNVRIVGCAVCGESKKAQVKEQLARKIEWLPSSAQALGIRNHYRHPEEH 300
              |||||||||||||||||||||||| |||||:|||||||||||||||||||||||||||||
Db        373 PLGAEWAEKADGNVRIVGCAVCGEFKKAQVQEQLARKIEWLPSSAQALGIRNHYRHPEEH 432

Qy        301 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        433 GSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTA 492

Qy        361 NLNRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGGAAK 420
              |||| |||||||||||||||||||||||:|||||||||| ||||||||||||||||||
Db        493 NLNRHAGKRYPFPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTGAGKPVDVIITGGGAAK 552

Qy        421 VAEALPPAFLAENTVRVADNLVIHGLLNLIAAEGGESEH 459
              |||||||||||||||||||||||:|||||:||||| | ||
Db        553 VAEALPPAFLAENTVRVADNLVIYGLLNMIAAEGREYEH 591


RESULT 3
I40327
baf protein - Bordetella pertussis
C;Species: Bordetella pertussis
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 08-Oct-1999
C;Accession: I40327; S70669
R;DeShazer, D.; Wood, G.E.; Friedman, R.L.
J. Bacteriol. 177, 3801-3807, 1995
A;Title: Identification of a Bordetella pertussis regulatory factor required for tran
A;Reference number: I40327; MUID:95325323; PMID:7601846
A;Accession: I40327
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-267 <RES>
A;Cross-references: EMBL:U12020; NID:g687228; PIDN:AAA75361.1; PID:g687229
R;Allen, A.; Maskell, D.
Mol. Microbiol. 19, 37-52, 1996
A;Title: The identification, cloning and mutagenesis of a genetic locus required for
A;Reference number: S70669; MUID:96419162; PMID:8821935
A;Accession: S70669
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 239-267 <ALL>
A;Cross-references: EMBL:X90711; NID:g992967; PIDN:CAA62242.1; PID:g992968
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995
C;Genetics:
A;Gene: baf

  Query Match             10.8%;  Score 258.5;  DB 2;  Length 267;
  Best Local Similarity   31.8%;  Pred. No. 4.3e-12;
  Matches   84;  Conservative   43;  Mismatches  116;  Indels   21;  Gaps    8;

Qy        209 LLLEGGNSRLKWAWVE-NGTFATVGSAPYR----DLSPLGAEWAEKADGNVRIVGCAVCG 263
              :|:: ||||||  | : :   |    ||     ||  ||   |       | :|  | |
Db          3 ILIDSGNSRLKVGWFDPDAPQAAREPAPVAFDNLDLDALGRWLATLPRRPQRALGVNVAG 62

Qy        264 ESKKAQVKEQLAR---KIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALG--SRRFS-RN 317
              ::   :   |      | || :   |:|:|| ||:|:: |:|||   :|  :|: |
Db         63 LARGEAIAATLRAGGCDIRWLRAQPLAMGLRNGYRNPDQLGADRWACMVGVLARQPSVHP 122

Qy        318 ACVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTG 377
              :| | ||| |:| :  |  : || |:||  :|: :||  ||:|    |  :|  |
Db        123 PLLVASFGTATTLDTIGPDNVFPGGLILPGPAMMRGALAYGTAHLPLADGLVADYPIDTH 182

Qy        378 NAVASGMMDAVCGSIMMMHGRLKEKNGAGKPVDVIITGGG-------AAKVAEALPPAFL 430
              |:|||:  |  |:|:     |  :   |:  :: : |||       | ::      ||
Db        183 QAIASGIAAAQAGAIVRQW--LAGRQRYGQAPEIYVAGGGWPEVRQEAERLLAVTGAAFG 240

Qy        431 AENTVRVADNLVIHGLLNLIAAEG 454
              |       |: |: ||  | ||:|
Db        241 ATPQPTYLDSPVLDGLAAL-AAQG 263


RESULT 4
A83112
BirA bifunctional protein PA4280 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: A83112
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: A83112
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07668.1; GSPDB:GN
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: birA; PA4280
C;Superfamily: birA bifunctional protein

  Query Match              9.8%;  Score 233.5;  DB 2;  Length 312;
  Best Local Similarity   36.9%;  Pred. No. 4.1e-10;
  Matches   72;  Conservative   23;  Mismatches   89;  Indels   11;  Gaps    5;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGL-ETQIKWPNDLVVGRDKLGGI 60
```

```
Db        121 GENLYYSLAITVRGGARELEGLSLVVGLAVLRALESLGVQDTGLKWPNDLLQGGKKIAGI 180

Qy         61 LIETVRAGGKTV----AVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLL 116

Db        181 LLE---LSGDPADLCHVVIGIGINVNMKTATDIGQPWTSLREALGEL--VDRSRLLSALN 235

Qy        117 AELGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLET 176

Db        236 RQLADYLDRHSREGFAASREEWESCNLWQGAQVILSTALSSI-EGRVLGVDSRGALRLEV 294

Qy        177 AEGEQTVVSGEISLR 191

Db        295 DGVEQGFSGGELSLR 309


RESULT 5
H83111
ypothetical protein PA4279 [imported] - Pseudomonas aeruginosa (strain PAO1)
  Species: Pseudomonas aeruginosa
;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83111
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83111
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-248 <STO>
A;Cross-references: GB:AE004843; GB:AE004091; NID:g9950489; PIDN:AAG07667.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA4279

  Query Match             9.8%;  Score 232.5;  DB 2;  Length 248;
  Best Local Similarity   27.2%;  Pred. No. 3.6e-10;
  Matches   70;  Conservative   41;  Mismatches  109;  Indels   37;  Gaps    7;

Qy        209 LLLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKA 268

Db          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Qy        269 QVKEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 321

Db         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Qy        322 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRPAGKRYPFPT 374

Db        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Qy        375 TTGNAVASGMMDAVCGSIMMMHGRLKEKNG-----AGKPVDVIITGGGAAKVAEALPPAF 429

Db        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

Qy        430 LAENTVRVADNLVIHGL 446

Db        231 ------RIMPDLVFVGL 241


RESULT 6
E82337
birA bifunctional protein VC0319 [imported] - Vibrio cholerae (strain N16961 serogroup O
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: E82337
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: E82337
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-320 <HEI>
A;Cross-references: GB:AE004120; GB:AE003852; NID:g9654727; PIDN:AAF93492.1; GSPDB:G
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VC0319
A;Map position: 1
C;Superfamily: birA bifunctional protein

  Query Match             9.0%;  Score 214.5;  DB 2;  Length 320;
  Best Local Similarity   31.8%;  Pred. No. 1.2e-08;
  Matches   62;  Conservative   27;  Mismatches  103;  Indels    3;  Gaps    3;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGGI 60

Db        127 GANLYLSIYWRLDAGMAAAMGLSLVVGVAIVEALEAMGVDGVKLKWPNDLYYQDKKLAGI 186

Qy         61 LIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 119

Db        187 LVEMSGQAGAAAHLVIGMGINLAMRDNEGNIDQPWISLAEVTGQSRIDRNALAINLIAAL 246

Qy        120 GAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEG 179

Db        247 DRTLRQYEISGMQNFVERWNRWDNFIGRPVKLLM-GANEVRGIERGIDEHGGVLLETEEG 305

Qy        180 EQTVVSGEISLRPDN 194

Db        306 LKSFIGGEISLRKND 320


RESULT 7
BVECBF
biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) - Escherichia coli (strain K-12
N;Contains: biotin operon repressor; biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.
C;Species: Escherichia coli
C;Date: 31-Mar-1988 #sequence_revision 31-Dec-1988 #text_change 03-Jun-2002
C;Accession: B24029; B45727; D65204; A24989
R;Howard, P.K.; Shaw, J.; Otsuka, A.J.
Gene 35, 321-331, 1985
A;Title: Nucleotide sequence of the birA gene encoding the biotin operon repressor a
A;Reference number: A91525; MUID:86006282; PMID:3899863
A;Accession: B24029
A;Molecule type: DNA
A;Residues: 1-321 <HOW>
A;Cross-references: GB:M10123; NID:g145430; PIDN:AAA23520.1; PID:g145432
R;Buoncristiani, M.R.; Howard, P.K.; Otsuka, A.J.
Gene 44, 255-261, 1986
A;Title: DNA-binding and enzymatic domains of the bifunctional biotin operon repress
A;Reference number: A24989; MUID:87055242; PMID:3536662
A;Contents: annotation; domain information
A;Note: this paper reports mutational studies done to understand DNA-binding and enz
R;Song, W.J.; Jackowski, S.
J. Bacteriol. 174, 6411-6417, 1992
A;Title: Cloning, sequencing, and expression of the pantothenate kinase (coaA) gene
A;Reference number: A45727; MUID:93015690; PMID:1328157
A;Accession: B45727
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 226-321 <SON>
A;Cross-references: GB:M90071
A;Note: residues 246-247 are correct to paper
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: D65204
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-321 <BLAT>
```

```
A;Cross-references: GB:AE000471; GB:U00096; NID:g1790404; PIDN:AAC76951.1; PID:g1790408;
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: birA
A;Map position: 89 min
C;Function: <ENZ>
A;Description: EC 6.3.4.15 [validated; MUID:99400973]; biotin-[acetyl-CoA-carboxylase] l
otin-accepting proteins
C;Function: <REP>
A;Description: acts as a biotin operon repressor [validated; MUID:99400973]; the reactic
orepressor
A;Note: the switch of birA from a biotin ligase to a repressor reflects competition betw
C;Superfamily: birA bifunctional protein
C;Keywords: ATP; biotin metabolism; DNA binding; ligase; repressor; transcription regula
F;1-26/Domain: DNA binding #status predicted <DNA>
F;83-119,189-235/Domain: biotin activation #status predicted <BIO>

  Query Match             8.8%;  Score 210;  DB 1;  Length 321;
  Best Local Similarity   33.7%;  Pred. No. 2.6e-08;
  Matches   65;  Conservative   29;  Mismatches   93;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| ||| : : || :   |           | |   |   |: |
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAGINLDRNTLAAMLIRE 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| ||:|: :|  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLAPYLSRWEKLDNFINRPVKLI-IGDKEIFGISRGIDKQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 IIKPWMGGEISLR 317


RESULT 8
D91241
hypothetical protein ECs4900 [imported] - Escherichia coli (strain O157:H7, substrain RI
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: D91241
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: D91241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB38323.1; PID:g13364376; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4900
C;Superfamily: birA bifunctional protein

  Query Match             8.8%;  Score 210;  DB 2;  Length 321;
  Best Local Similarity   33.7%;  Pred. No. 2.6e-08;
  Matches   65;  Conservative   29;  Mismatches   93;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| ||| : : || :   |           | |   |   |: |
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAGINLDRNTLAAMLIRE 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| ||:|: :|  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLAPYLSRWEKLDNFINRPVKLI-IGDKEIFGISRGIDKQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 IIKPWMGGEISLR 317


RESULT 9
A86089
hypothetical protein birA [imported] - Escherichia coli (strain O157:H7, substrain El
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: A86089
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Ma
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apod
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: A86089
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-321 <STO>
A;Cross-references: GB:AE005174; NID:g12518899; PIDN:AAG59173.1; GSPDB:GN00145; UWGP
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: birA
C;Superfamily: birA bifunctional protein

  Query Match             8.8%;  Score 210;  DB 2;  Length 321;
  Best Local Similarity   33.7%;  Pred. No. 2.6e-08;
  Matches   65;  Conservative   29;  Mismatches   93;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| ||| : : || :   |           | |   |   |: |
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAGINLDRNTLAAMLIRE 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| ||:|: :|  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLAPYLSRWEKLDNFINRPVKLI-IGDKEIFGISRGIDKQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 IIKPWMGGEISLR 317


RESULT 10
A82637
conserved hypothetical protein XF1795 [imported] - Xylella fastidiosa (strain 9a5c)
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: A82637
R;anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Se
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82637
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-242 <SIM>
A;Cross-references: GB:AE004001; GB:AE003849; NID:g9106864; PIDN:AAF84603.1; GSPDB:GI
A;Experimental source: strain 9a5c
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrei
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fi
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins

  Query Match             8.8%;  Score 210;  DB 2;  Length 321;
  Best Local Similarity   33.7%;  Pred. No. 2.6e-08;
  Matches   65;  Conservative   29;  Mismatches   93;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |  |  |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMFWRLEQGPAAAIG-LSLVIGIVMAEVLRKLGADKVRVKWPNDLYLQDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:| | : |     |:| ||| : : || :   |           | |   |   |: |
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEESVVNQGWITLQEAGINLDRNTLAAMLIRE 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| ||:|: :|  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLAPYLSRWEKLDNFINRPVKLI-IGDKEIFGISRGIDKQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 IIKPWMGGEISLR 317
```

```
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuhako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1795

  Query Match             8.4%;  Score 199.5;  DB 2;  Length 242;
  Best Local Similarity   30.3%;  Pred. No. 1.1e-07;
  Matches   74;  Conservative   31;  Mismatches  116;  Indels   23;  Gaps    8;

Qy        210 LLEGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGESKKAQ 269
              | : |||| | | :  |    |   ||   :      |   | |  : :|   :
Db          5 LFDLGNSRFKCASLREGVIGPVTVLPYLTETMDAFALQELPRGRVAYL-ASVAAPAITTH  63

Qy        270 VKEQLA---RKIEWLPSSAQALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGT 326
              | | |     :::   : |   |:|  | |||  | ||:   |||  :     :||  ||
Db         64 VLEVLKIHFEQVQVAATVAACAGVRIAYAHPERFGVDRFLALLGS--YGEGNVLVVGVGT 121

Qy        327 AVTVDALTDDGHYLGGTIMPGFHLMKESLAVRTANLNRPAGKRYPFPTTTGNAVASGMMD 386
              |:|:| |  :| :||| |     ||:::|  |   |    |    |   | :|: ||
Db        122 ALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGNYLEFAEDTEDALVSGCNG 181

Qy        387 AVCG----SIMMMHGRLKEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLV 442
              |       |:   | || :       | ::: |||   ||  ||  :| :   |    ||
Db        182 AAVALIERSLYEAHQRLDQS------VRLLLHGGG---VASLLP--WLGDVVHR--PTLV 228

Qy        443 IHGL 446
              : ||
Db        229 LDGL 232


RESULT 11
AG0457
biotin-[acetyl-CoA-carboxylase] ligase (EC 6.3.4.15) [imported] - Yersinia pestis (strai
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C;Accession: AG0457
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AG0457
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-319 <KUR>
A;Cross-references: GB:AL590842; PIDN:CAC93227.1; PID:g15981675; GSPDB:GN00175
C;Genetics:
A;Gene: birA
C;Superfamily: birA bifunctional protein
C;Keywords: ligase

  Query Match             8.1%;  Score 194;  DB 2;  Length 319;
  Best Local Similarity   29.7%;  Pred. No. 4.1e-07;
  Matches   57;  Conservative   30;  Mismatches  101;  Indels    4;  Gaps    4;

Qy          2 GECLMFSFGWAFDRPQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGGI 60
              |  |  |  |  ::       || |  :     |  || |  ::|||||  :   || ||
Db        128 GANLYLSMYWRLEQGPAAAMGLSLVVGIVMAEVLQNLGAEKVRVKWPNDLYLNDKKLAGI 187

Qy         61 LIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAEL 119
              |:| | :|     |:| |||  :     || : |           |   |   :|:||
Db        188 LVELTGKTGDAAQLVIGAGINLTMRGSETNAIN-QDWINLQDAGVTIDRNKLTAEILSEL 246

Qy        120 GAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAEG 179
                : ::  || : ||: ::  :    : | |:   : :  |  :|:| :| | ||
Db        247 RLAVVKFENEGLSAFLSRWQKMDNYLNRPVKLIIGNQEIF-GIARGIDQQGALLLEQDGN 305

Qy        180 EQTVVSGEISLR 191
                : : ||||||
Db        306 IKPYIGGEISLR 317


RESULT 12
AH1684
transcription regulator and biotin acetyl-CoA-carboxylase synthetase homolog birA [i
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AH1684
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloe
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1684
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-325 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97248.1; PID:g16414519; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: birA

  Query Match             8.0%;  Score 190;  DB 2;  Length 325;
  Best Local Similarity   31.3%;  Pred. No. 8.5e-07;
  Matches   62;  Conservative   37;  Mismatches   55;  Indels   44;  Gaps   12;

Qy         16 PQYELGSLSPVAALACRRAL-GCLGLETQIKWPNDLVVGRDKLGGILIETVRAGGKTV-- 72
              ||:       :|:||   |:     || :||||||: :|: |: |:| | ::|  :|:
Db        151 PQFTF-----IASLAITEAIENITKLEPKIKWPNDIYIGKRKICGVLTE-MQAEAETIHA 204

Qy         73 AVVGIGINF---VLPKEV-ENAASVQ-SLFQTASRRGNADAAVLLETLLAELGAVLEQYA 127
                ::|:|||      |:|: | |:|::  | :: ||:       ||: :|  |    | :
Db        205 VIIGMGINVNQQEFPEEIKEKASSLKLELGESVSRK------ALLQAILTSLEKYYELFL 258

Qy        128 EEGFAPF----------LNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETA 177
              ::|||              |  ||:   ||             | |||:   ||| |: :
Db        259 DKGFAPIKLLWETKAIPFGEKLTASTTKGK----------IHGQVKGISDEGVLLLQDS 307

Qy        178 EGE-QTVVSGEISLRPDN 194
                ||  :: | :| |  ||
Db        308 LGEIHSIYSADILL--DN 323


RESULT 13
AG0934
bifunctional protein, biotin operon repressor and biotin-[acetyl-CoA carboxylase] sy
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C;Accession: AG0934
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churc
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Far
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica s
A;Reference number: AB0502; PMID:11677608
A;Accession: AG0934
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-320 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD09497.1; PID:g16504614; GSPDB:GN00176
C;Genetics:
A;Gene: STY3741
C;Superfamily: birA bifunctional protein

  Query Match             7.8%;  Score 187;  DB 2;  Length 320;
```

```
  Best Local Similarity   31.6%;  Pred. No. 1.4e-06;
  Matches   61;  Conservative   32;  Mismatches   94;  Indels    6;  Gaps    6;

Qy          2 GECLMFSFGWAFDR-PQYELGSLSPVAALACRRALGCLGLE-TQIKWPNDLVVGRDKLGG 59
              |  |   |   |  :: |   :| || |  :     |  || :  ::|||||| :   || |
Db        128 GANLYLSMYWRLEQGPAAAIG-LSLVIGIVMAEVLRELGADKVRVKWPNDLYLLDRKLAG 186

Qy         60 ILIE-TVRAGGKTVAVVGIGINFVLPKEVENAASVQSLFQTASRRGNADAAVLLETLLAE 118
              ||:|  |  : |     |:| |||  : : ||     |            |  :|   |: :
Db        187 ILVELTGKTGDAAQIVIGAGINMAM-RRVEEDVINQGWITLQEAGITLDRNMLAAKLIYK 245

Qy        119 LGAVLEQYAEEGFAPFLNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETAE 178
              | | || : :|| :|:|: ::  :    : | |:  |:    |  :|:| :| | ||
Db        246 LRAALELFEQEGLSPYLSRWKKLDNFIDRPVKLI-IGDKEIFGISRGIDTQGALLLEQDG 304

Qy        179 GEQTVVSGEISLR 191
                :  : ||||||
Db        305 VIKPWMGGEISLR 317


RESULT 14
AH1312
transcription regulator and biotin acetyl-CoA-carboxylase synthetase homolog birA [impor
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AH1312
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1312
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-325 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99982.1; PID:g16411357; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: birA

  Query Match             7.8%;  Score 187;  DB 2;  Length 325;
  Best Local Similarity   30.8%;  Pred. No. 1.4e-06;
  Matches   61;  Conservative   35;  Mismatches   58;  Indels   44;  Gaps   11;

Qy         16 PQYELGSLSPVAALACRRAL-GCLGLETQIKWPNDLVVGRDKLGGILIETVRAGGKTV-- 72
              ||:       :|:||   |:     || :||||||: :|: |: |:| | ::|  :|:
Db        151 PQFTF-----IASLAITEAIENITKLEPKIKWPNDIYIGKRKICGVLTE-MQAEAETIHA 204

Qy         73 AVVGIGINF---VLPKEVENAAS--VQSLFQTASRRGNADAAVLLETLLAELGAVLEQYA 127
               ::|:|||      |:|::: ||     | :: ||:        ||: :|  |    | :
Db        205 VIIGMGINVNQQEFPEEIKDKASSLKLELGESVSRK------ALLQEILTSLEKYYELFL 258

Qy        128 EEGFAPF----------LNEYETANRDHGKAVLLLRDGETVCEGTVKGVDGRGVLHLETA 177
              ::||||           |  ||:   ||            | |||:   ||| |:
Db        259 DKGFAPIKLLWETKAIPFGEKLTASTTKGK----------IHGQVKGISDEGVLLLQDT 307

Qy        178 EGE-QTVVSGEISLRPDN 194
               ||  :: | :| |  ||
Db        308 LGEVHSIYSADILL--DN 323


RESULT 15
B87489
transcription activator, probable Baf family [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: B87489
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: B87489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <STO>
A;Cross-references: GB:AE005673; NID:g13423392; PIDN:AAK23910.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC1935
C;Superfamily: Streptomyces coelicolor hypothetical protein SCE94.31c

  Query Match             7.8%;  Score 185;  DB 2;  Length 261;
  Best Local Similarity   34.1%;  Pred. No. 1.5e-06;
  Matches   58;  Conservative   23;  Mismatches   69;  Indels   20;  Gaps    6;

Qy        288 LGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVVVSCGTAVTVDALTDDGHYLGGTIMPG 347
              |||       | | |:||  ||:|:         ||:  ||| | | :  || : || | ||
Db         95 LGIDVRIEKPSEAGADRLVNAIGAAMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPG 154

Qy        348 FHLMKESLAVRTANL-----NRPAGKRYPFPTTTGNAVASGMMDAVCGSIMMMHG---RL 399
               :|  ::|    | |       |||| |    | | :|: ||:     | | :: |   |:
Db        155 INLSMQALHEAAAKLPRIAIQRPAGNRI-VGTDTVSAMQSGVF---WGYISLIEGLVARI 210

Qy        400 KEKNGAGKPVDVIITGGGAAKVAEALPPAFLAENTVRVADNLVIHGLLNL 449
              | :   |:|: || ||| |:    |       ::      :| | ||| :
Db        211 KAER--GEPMTVIATGGVASLFEGA------TDSIDHFDSDLTIRGLLEI 252


Search completed: June 24, 2003, 22:03:21
Job time : 17.7781 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 24, 2003, 21:35:06 ; Search time 20.2269 Seconds
(without alignments)
2526.317 Million cell updates/sec

Title: US-09-813-453A-20
Perfect score: 1248
Sequence: 1 MILELDCGNSLIKWRVIEGA.....GARIMPDLVFVGLALACPIE 248

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1248	100.0	248	16	Q9HWC1	Q9hwc1 pseudomonas
2	234.5	18.8	592	16	Q9JXF1	Q9jxf1 neisseria m
3	231.5	18.5	295	16	Q8Y2M4	Q8y2m4 ralstonia s
4	231.5	18.5	592	16	Q9JWI7	Q9jwi7 neisseria m
5	193	15.5	242	16	Q9PCI4	Q9pci4 xylella fas
6	176.5	14.1	273	16	Q97EB4	Q97eb4 clostridium
7	172.5	13.8	276	16	Q8YQD7	Q8yqd7 anabaena sp
8	160	12.8	261	16	Q9A6Z1	Q9a6z1 caulobacter
9	159.5	12.8	223	16	O25533	O25533 helicobacte
10	158.5	12.7	223	16	Q9ZKY6	Q9zky6 helicobacte
11	148.872	11.9	272	16	O06282	O06282 mycobacteri
12	145	11.6	246	16	Q9WZY5	Q9wzy5 thermotoga
13	144	11.5	255	16	Q8R7M2	Q8r7m2 thermoanaer
14	140	11.2	259	16	Q8XHL5	Q8xhl5 clostridium
15	138.5	11.1	265	16	Q9X8N6	Q9x8n6 streptomyce
16	137.372	11.0	274	16	Q9CD56	Q9cd56 mycobacteri
17	136	10.9	273	16	O83446	O83446 treponema p
18	131.5	10.5	259	16	Q8YAC5	Q8yac5 listeria mo
19	126.5	10.1	259	16	Q92F54	Q92f54 listeria in
20	121	9.7	254	16	Q9KGH5	Q9kgh5 bacillus ha
21	120.5	9.7	209	16	Q9PIA9	Q9pia9 campylobact
22	120	9.6	256	16	Q8RFE4	Q8rfe4 fusobacteri
23	117.5	9.4	258	2	Q9F985	Q9f985 bacillus st
24	107	8.6	212	2	O32514	O32514 desulfovibr
25	107	8.6	257	16	P74045	P74045 synechocyst
26	101.5	8.1	451	10	Q9LXJ6	Q9lxj6 arabidopsis
27	101	8.1	522	16	Q8XS12	Q8xs12 ralstonia s
28	100.5	8.1	518	16	Q98BS3	Q98bs3 rhizobium l
29	98.5	7.9	229	16	O67753	O67753 aquifex aeo
30	97.5	7.8	238	16	Q9L007	Q9l007 streptomyce
31	97	7.8	728	16	Q9RZ81	Q9rz81 deinococcus
32	96.5	7.7	509	16	Q92AP3	Q92ap3 listeria in
33	96.5	7.7	977	16	Q9RKT0	Q9rkt0 streptomyce
34	93.5	7.5	481	16	Q92S44	Q92s44 rhizobium m
35	92	7.4	224	16	Q98Q93	Q98q93 mycoplasma
36	92	7.4	349	16	Q8XSJ1	Q8xsj1 ralstonia s
37	92	7.4	1599	11	Q99NH0	Q99nh0 mus musculu
38	91.5	7.3	822	16	Q9RX95	Q9rx95 deinococcus
39	90.5	7.3	600	16	Q8Y088	Q8y088 ralstonia s
40	90.5	7.3	841	16	Q9ACU8	Q9acu8 streptomyce
41	90.5	7.3	940	16	Q9A6P6	Q9a6p6 caulobacter
42	90	7.2	1627	5	Q962Q0	Q962q0 giardia lam
43	89.5	7.2	509	16	Q8Y6C5	Q8y6c5 listeria mo
44	89	7.1	419	16	P74122	P74122 synechocyst
45	89	7.1	908	16	Q8ZD99	Q8zd99 yersinia pe

ALIGNMENTS

RESULT 1
Q9HWC1
ID Q9HWC1 PRELIMINARY; PRT; 248 AA.
AC Q9HWC1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein PA4279.
GN PA4279.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrener P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004843; AAG07667.1; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR001230; Prenyl_site.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 248 AA; 26758 MW; 609D37480899DA3B CRC64;

Query Match 100.0%; Score 1248; DB 16; Length 248;
Best Local Similarity 100.0%; Pred. No. 3.4e-94;
Matches 248; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60

```
Db          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60

Qy         61 TSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVI 120

Db         61 TSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLVI 120

Qy        121 DLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQAT 180

Db        121 DLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQAT 180

Qy        181 AEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVG 240

Db        181 AEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVFVG 240

Qy        241 LALACPIE 248

Db        241 LALACPIE 248
```

RESULT 2
Q9JXF1
ID Q9JXF1 PRELIMINARY; PRT; 592 AA.
AC Q9JXF1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BirA protein/Bvg accessory factor.
GN NMB2075.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
DR EMBL; AE002557; AAF42394.1; -.
DR HSSP; P06709; 1BIA.
DR TIGR; NMB2075; -.
DR InterPro; IPR004619; Baf.
DR InterPro; IPR004408; BirA_ligase.
DR InterPro; IPR003142; BPL_C.
DR InterPro; IPR004143; BPL_LipA_LipB.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF02237; BPL_C; 1.
DR Pfam; PF03099; BPL_LipA_LipB; 1.
DR Pfam; PF03309; Bvg_acc_factor; 1.
DR TIGRFAMs; TIGR00121; birA_ligase; 1.
DR PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 592 AA; 64701 MW; 9E27421DA2B41DE3 CRC64;

Query Match 18.8%; Score 234.5; DB 16; Length 592;
Best Local Similarity 27.6%; Pred. No. 5.5e-11;
Matches 71; Conservative 40; Mismatches 109; Indels 37; Gaps 7;

```
Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA---LPVRACRLVSVRSEQ 59

Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRDLSPLGAEWAEKADGNVRIVGCAVCGEFKKA 400

Qy         60 ETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACLV 119

Db        401 QVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNACVV 453

Qy        120 IDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP--- 176

Db        454 VSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYPFPT 506

Qy        177 --GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGA- 230

Db        507 TTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALPPAF 561

Qy        231 ------RIMPDLVFVGL 241

Db        562 LAENTVRVADNLVIYGL 578
```

RESULT 3
Q8Y2M4
ID Q8Y2M4 PRELIMINARY; PRT; 295 AA.
AC Q8Y2M4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative transcription regulation accessory factor transcription
DE regulator protein.
GN RSC0311 OR RS03278.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646058; CAD13839.1; -.
DR InterPro; IPR004619; Baf.
DR Pfam; PF03309; Bvg_acc_factor; 1.
KW Complete proteome.
SQ SEQUENCE 295 AA; 30990 MW; 133074A7764BFCBB CRC64;

Query Match 18.5%; Score 231.5; DB 16; Length 295;
Best Local Similarity 31.0%; Pred. No. 4e-11;
Matches 86; Conservative 36; Mismatches 106; Indels 49; Gaps 13;

```
Qy          2 ILELDCGNSLIK--WRVIEGAARSVA-GGL------AESDDALVEQLTSQQALPVRACR- 51

Db         13 LLLIDAGNTRIKWAWTAADVAPPAVAPGGTPWQHAQARPHDQLAELVEDW-----RDCHA 67

Qy         52 ----------LVSVRSEQETSQLVARLEQLFPVSAL-VASSGKQLAGVRNGYLDYQRLGL 100

Db         68 GAGMAPPDVWISVVAGPALRDALCARIARVFDGARLRIVASEAAAAGLRNGYRDPAQLGT 127

Qy        101 DRWLALVAAHHL-AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRR 159

Db        128 DRWVGAVGARHAWPDTALLLVTAGTATTLDIVAPDGRFAGGLILPGLTLMMRALSRNTAQ 187

Qy        160 I---------RYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQY-AMACELLGP 209

Db        188 LPEIDIGYLAARDDAQAPADVPSW--ADNTQDAIALGCVTAQAGAIAQTWQALQAQYPGP 245

Qy        210 DCEIFLTGGDAELVRDELAGARIMP-----DLVFVGL 241

Db        246 -YRCVLSGG----ARAALAPHLRMPFQMHDNLVLLGL 277
```

RESULT 4

```
Q9JWI7
ID   Q9JWI7      PRELIMINARY;      PRT;   592 AA.
AC   Q9JWI7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BirA bifunctional protein (EC 6.3.4.15).
GN   BIRA OR NMA0357.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162752; CAB83659.1; -.
DR   HSSP; P06709; 1BIA.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR004408; BirA_ligase.
DR   InterPro; IPR003142; BPL_C.
DR   InterPro; IPR004143; BPL_LipA_LipB.
DR   InterPro; IPR001525; C5_DNA_meth.
DR   Pfam; PF02237; BPL_C; 1.
DR   Pfam; PF03099; BPL_LipA_LipB; 1.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00121; birA_ligase; 1.
DR   PROSITE; PS00095; C5_MTASE_2; UNKNOWN_1.
KW   Ligase; Complete proteome.
SQ   SEQUENCE   592 AA;  64468 MW;  7ED2E0CD3B31C630 CRC64;

  Query Match             18.5%;  Score 231.5;  DB 16;  Length 592;
  Best Local Similarity   27.7%;  Pred. No. 9.7e-11;
  Matches   72;  Conservative   38;  Mismatches  107;  Indels   43;  Gaps    8;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQA------LPVRACRLVSVR 56
              | || ||| :|| :|    :  |     |   :  | :: |      : :  | :     .
Db        341 LLLDGGNSRLKWAWVENGTFATVGSAPYRD---LSPLGAEWAEKVDGNVRIVGCAVCGEF 397

Qy         57 SEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKA 116
                :  : :|  :  | ||       |  |:|| |   :  | |||   : :   :: |
Db        398 KKAQVQEQLARKIEWLPSSA-------QALGIRNHYRHPEEHGSDRWFNALGSRRFSRNA 450

Qy        117 CLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQP 176
              |:|:  ||||| | :  || :||| | ||  ||:  |   |       |   |      |
Db        451 CVVVSCGTAVTVDALTDDGHYLGGTIMPGFHLMKESLAVRT-------ANLNRHAGKRYP 503

Qy        177 -----GQATAEAVER---GCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELA 228
                  | | |  :      | ::|: | ::|:        |   :: :||| |  | : |
Db        504 FPTTTGNAVASGMMDAVCGSVMMMHGRLKEKTG-----AGKPVDVIITGGGAAKVAEALP 558

Qy        229 GA-------RIMPDLVFVGL 241
               |       |:  :||  ||
Db        559 PAFLAENTVRVADNLVIHGL 578


RESULT 5
Q9PCI4
ID   Q9PCI4      PRELIMINARY;      PRT;   242 AA.
AC   Q9PCI4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Xf1795.
GN   XF1795.
OS   Xylella fastidiosa.
OC   Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC   Xylella.
OX   NCBI_TaxID=2371;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=9A5C;
RX   MEDLINE=20365717; PubMed=10910347;
RA   Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA   Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA   Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA   Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA   Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA   Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA   Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA   Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA   Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA   Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA   Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA   Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA   Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA   Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA   Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA   Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA   Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA   de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA   Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA   Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA   de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA   da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA   da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA   de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsuhako M.H.,
RA   Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA   Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT   "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL   Nature 406:151-159(2000).
DR   EMBL; AE004001; AAF84603.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   242 AA;  25658 MW;  EDC5DE7C0449C1E9 CRC64;

  Query Match             15.5%;  Score 193;  DB 16;  Length 242;
  Best Local Similarity   31.0%;  Pred. No. 4.4e-08;
  Matches   76;  Conservative   34;  Mismatches  111;  Indels   24;  Gaps    8;

Qy          6 DCGNSLIKWRVI-EGAAR--SVAGGLAESDDALVEQLTSQQALP-VRACRLVSVRSEQET 61
              | |||  |   : ||     :|   | |: ||      : | ||  |   | || :   |
Db          7 DLGNSRFKCASLREGVIGPVTVLPYLTETMDAF-----ALQELPRGRVAYLASVAAPAIT 61

Qy         62 SQLVARLEQLF---PVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHLAKKACL 118
              : ::  |:  |    |:| ||:     ||||  |   :| |:||:|||: ::   :  |
Db         62 THVLEVLKIHFEQVQVAATVAA----CAGVRIAYAHPERFGVDRFLALLGSY--GEGNVL 115

Qy        119 VIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQ 178
              |: :|||:| ||:||:| |||| |    ||||  |     ::                |: :
Db        116 VVGVGTALTIDLLAANGCHLGGRISASPTLMRQALHARAEQLPLSGGN------YLEFAE 169

Qy        179 ATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLTGGDAELVRDELAGARIMPDLVF 238
              | :|:  ||       :      | : |    : | ||    :  |       | ||
Db        170 DTEDALVSGCNGAAVALIERSLYEAHQRLDQSVRLLLHGGGVASLLPWLGDVVHRPTLVL 229

Qy        239 VGLAL 243
               |||:
Db        230 DGLAI 234


RESULT 6
Q97EB4
ID   Q97EB4      PRELIMINARY;      PRT;   273 AA.
AC   Q97EB4;
DT   01-OCT-2001 (TrEMBLrel. 18, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
```

```
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Predicted transcriptional regulator, homolog of Bvg accessory
DE   factor.
GN   CAC3200.
OS   Clostridium acetobutylicum.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1488;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX   MEDLINE=21359325; PubMed=11466286;
RA   Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA   Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA   Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA   Bennett G.N., Koonin E.V., Smith D.R.;
RT   "Genome sequence and comparative analysis of the solvent-producing
RT   bacterium Clostridium acetobutylicum.";
RL   J. Bacteriol. 183:4823-4838(2001).
DR   EMBL; AE007815; AAK81136.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   273 AA;  30331 MW;  DE7B4D5923E72842 CRC64;

  Query Match             14.1%;  Score 176.5;  DB 16;  Length 273;
  Best Local Similarity   26.5%;  Pred. No. 1.1e-06;
  Matches   71;  Conservative   43;  Mismatches  113;  Indels   41;  Gaps    9;

Qy          1 MILELDCGNSLI------------KWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVR 48
              :|| || ||: |             :||:     ||     |:     |  |  |  |
Db         12 VILVLDVGNTNIVLGIYNDTKLTAEWRLSTDVLRS-----ADEYGIQVMNLFQQDKLDPT 66

Qy         49 ACRLVSVRSEQETSQLVARLEQL----FPVSALVASSGKQLAGVRNGYLDYQRLGLDRWL 104
                | : |      ::  || :    | :: ||   | :  |:   | : : :| || :
Db         67 LVEGVIISS--VVPNIMYSLEHMIRKYFKINPLVVGPGIK-TGINIKYDNPKEVGADRIV 123

Qy        105 ALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDD 164
              ||||  : |:: ::|| ||| |  ' |  :| :||| ||||: :   |       ::
Db        124 NAVAAHEIYKRSLIIIDFGTATTFCAVRENGDYLGGAICPGIKVSSEAL--------FEK 175

Qy        165 AEARRALASLQPGQA----TAEAVERGCLLMLRGFVREQYAMACELLGPDCE----IFLT 216
              |     :  ::|  |    |  ::: | :   | ||       | |  : |    :  |
Db        176 AAKLPRVELIKPAYAICKNTISSIQSGIVYGYIGQVRYIVERMKEELQEEGEKEPLVVAT 235

Qy        217 GGDAELVRDELAGARIM-PDLVFVGLAL 243
              || |:|: :|     :: | |   || :
Db        236 GGLAKLISEEAKNVDVINPFLTLEGLRI 263


RESULT 7
Q8YQD7
ID   Q8YQD7      PRELIMINARY;      PRT;   276 AA.
AC   Q8YQD7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein Alr3896.
GN   ALR3896.
OS   Anabaena sp. (strain PCC 7120).
OC   Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX   NCBI_TaxID=103690;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21595285; PubMed=11759840;
RA   Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA   Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA   Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA   Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA   Yasuda M., Tabata S.;
RT   "Complete genomic sequence of the filamentous nitrogen-fixing
RT   cyanobacterium Anabaena sp. strain PCC 7120.";
RL   DNA Res. 8:205-213(2001).
DR   EMBL; AP003594; BAB75595.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   276 AA;  30272 MW;  2AD04CA693E56F25 CRC64;

  Query Match             13.8%;  Score 172.5;  DB 16;  Length 276;
  Best Local Similarity   27.1%;  Pred. No. 2.4e-06;
  Matches   76;  Conservative   43;  Mismatches   94;  Indels   67;  Gaps   12;

Qy          3 LELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQETS 62
              | |: ||| : | :  | :          :::::|| : :          : | ||:
Db         15 LALEIGNSRLHWALFMGESLEFTWDTEYLPESVIQQLGNGET-------KLEVGSEE--- 64

Qy         63 QLVARLEQLF-----------PVSALVASSGKQLAGVRNGYLD--------------YQR 97
                 :::|            |:   :||   |   :   ||:              |
Db         65 ------KEIFFTFFPLPPAPCPLPLFIASVVPQQTVLWENYLNVRVITLDQIPLNNIYPT 118

Qy         98 LGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVH--LGGYICPGMTLMRSQLRT 155
              ||:|| |||  |         |||| |||:|    ||||    :|| | ||: |  : |
Db        119 LGIDRALALWGAGMSWGFPVLVIDAGTALT--FTAADGGKNLVGGAILPGVGLQFASLGQ 176

Qy        156 HTRRIRYDDAEARRALASLQPGQA--TAEAVERGCLLML----RGFVREQYAMACELLGP 209
              | ::   : |   |: || |  |  | ||:: | :  |    | | |  :     | |
Db        177 QTGQLPQVEME---AIKSLPPRFALNTTEAIQSGVIYTLIAGMRDFTEEWLS-----LFP 228

Qy        210 DCEIFLTGGDAELVRDEL------AGARIM--PDLVFVGL 241
              | :: : |||  |: : |        ||::  |:|:| |:
Db        229 DGKVAIKGGDRILLLNYLQALYPDLAARLIVEPNLIFWGM 268


RESULT 8
Q9A6Z1
ID   Q9A6Z1      PRELIMINARY;      PRT;   261 AA.
AC   Q9A6Z1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Transcriptional activator, putative, Baf family.
GN   CC1935.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005867; AAK23910.1; -.
DR   TIGR; CC1935; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   261 AA;  27965 MW;  C19E60D7B0714EF5 CRC64;

  Query Match             12.8%;  Score 160;  DB 16;  Length 261;
  Best Local Similarity   28.3%;  Pred. No. 2.4e-05;
  Matches   67;  Conservative   30;  Mismatches  116;  Indels   24;  Gaps    8;

Qy          1 MILELDCGNSLIKWRVIEGAA-----RSVAGGLAESDDALV--EQLTSQQALPVRACRLV 53
```

```
              |:| :: ||:   : : :||:     ||       :|: :|   || | | |  ||   |
Db          2 MLLAIEQGNTNTMFAIHDGASWVAQWRSATESTRTADEYVVWLSQLLSMQGLGFRAIDAV 61

Qy         54 SVRSEQETSQLVARLEQL----FPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAA 109
               : |     | :  |  |    | |  ||     :| |:          | || :  : |
Db         62 IISS--VVPQSIFNLRNLSRRYFNVEPLVIGENAKL-GIDVRIEKPSEAGADRLVNAIGA 118

Qy        110 HHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARR 169
                :     :||| ||| | |:|||||   || | ||: |    |  |    :      :|
Db        119 AMVYPGPLVVIDSGTATTFDIVAADGAFEGGIIAPGINLSMQAL--HEAAAKLPRIAIQR 176

Qy        170 ALASLQPGQATAEAVERGC----LLMLRGFVREQYAMACELLGPDCEIFLTGGDAEL 222
                  :  | |  |:: |     : :: | |    |      |    :  ||| | |
Db        177 PAGNRIVGTDTVSAMQSGVFWGYISLIEGLVARIKAER----GEPMTVIATGGVASL 229


RESULT 9
O25533
ID   O25533      PRELIMINARY;      PRT;   223 AA.
AC   O25533;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein HP0862.
GN   HP0862.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
DR   EMBL; AE000596; AAD07916.1; -.
DR   TIGR; HP0862; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   223 AA;  24655 MW;  C1D13E8BEC1EB9AF CRC64;

  Query Match             12.8%;  Score 159.5;  DB 16;  Length 223;
  Best Local Similarity   28.6%;  Pred. No. 2.1e-05;
  Matches   58;  Conservative   34;  Mismatches   60;  Indels   51;  Gaps    8;

Qy         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112
               :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| ::
Db         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYVGLGIDRQMACLAVNN- 97

Qy        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172
                 :|:| |:|:| ||:  :| |||| | ||:               :|  |  : |
Db         98 ----GVVVDAGSAITIDLI-KEGKHLGGCILPGLA------------QYIHAYKKSAKI 139

Qy        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221
                 || :|         | :||  | :| :          :|| : |  : :|:| ||||:
Db        140 LEQPFKALDSLEVLPKSTRDAVNYGMVLSV---------IACIQHLAKNQKIYLCGGDAK 190

Qy        222 LVRDELAGARIMPDLVFVGLALA 244
                :   |  :     ||| |: :|
Db        191 YLSAFLPHSVCKERLVFDGMEIA 213


RESULT 10
Q9ZKY6
ID   Q9ZKY6      PRELIMINARY;      PRT;   223 AA.
AC   Q9ZKY6;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   PUTATIVE.
GN   JHP0796.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
DR   EMBL; AE001509; AAD06372.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Complete proteome.
SQ   SEQUENCE   223 AA;  24754 MW;  B31377DA1AEF6E0E CRC64;

  Query Match             12.7%;  Score 158.5;  DB 16;  Length 223;
  Best Local Similarity   28.6%;  Pred. No. 2.6e-05;
  Matches   58;  Conservative   33;  Mismatches   61;  Indels   51;  Gaps    8;

Qy         53 VSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALVAAHHL 112
               :||  | | : |       :|       : | :||  :   ||  ||:|| :| :| :
Db         51 ISVNEENEKALL-----NCYP-------NAKNIAGFFHLETDYIGLGIDRQMACLAVVN- 97

Qy        113 AKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEARRALA 172
                 :::| |:|:| |||  :| |||| | ||:               :|  |  : |
Db         98 ----GVIVDAGSAITIDLV-KEGKHLGGCILPGLA------------QYVHAYKKSAKI 139

Qy        173 SLQPGQA----------TAEAVERGCLLMLRGFVREQYAMAC-ELLGPDCEIFLTGGDAE 221
                 || :|         | :||  | :| :          ::| : |  | :|:| ||||:
Db        140 LEQPFKALDSLEVLPKNTRDAVNYGMILSI---------ISCIQHLAKDQKIYLCGGDAK 190

Qy        222 LVRDELAGARIMPDLVFVGLALA 244
                :   |  :     ||| |: :|
Db        191 YLSAFLPHSVCKERLVFDGMEIA 213


RESULT 11
O06282
ID   O06282      PRELIMINARY;      PRT;   272 AA.
AC   O06282;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 29.3 kDa protein (Transcriptional activator, putative,
DE   Baf family).
GN   RV3600C OR MTCY07H7B.22 OR MT3706.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
```

```
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z95557; CAB08944.1; -.
DR   EMBL; AE007170; AAK48063.1; -.
DR   TIGR; MT3706; -.
DR   TubercuList; Rv3600c; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   272 AA;  29304 MW;  5D70E6EF0F09AC8B CRC64;

  Query Match           11.9%;  Score 148.872;  DB 16;  Length 272;
  Best Local Similarity  22.9%;  Pred. No. 0.0002;
  Matches   62;  Conservative   44;  Mismatches  103;  Indels   62;  Gaps    8;

Qy          1 MILELDCGNS----------------LIKWRVIEGAARSVAGGLAESDDALV----EQLT 40
              |:| :|  |:                : :|| |   :    |  || : | |:    |:||
Db          1 MLLAIDVRNTHTVVGLLSGMKEHAKVVQQWR-IRTESEVTADELALTIDGLIGEDSERLT 59

Qy         41 SQQALPVRACRLVSVRSEQETSQLVARLEQLFPVSALVASSGKQLAGVRNGYL----DYQ 96
                ||      |  ||           |:| :|    |       |||  |      : :
Db         60 GTAALSTVPSVLHEVR---------IMLDQYWPSVPHVLIE----PGVRTGIPLLVDNPK 106

Qy         97 RLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTH 156
              :| || :  :||:   :|| :|:| |:::  |:|:| |  ||| | ||: :
Db        107 EVGADRIVNCLAAYDRFRKAAIVVDFGSSICVDVVSAKGEFLGGAIAPGVQVSSDAAAAR 166

Qy        157 TRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLT 216
              :  :|  :    |::     |: | | :: | :    |                    |
Db        167 SAALRRVELARPRSVV----GKNTVECMQAGAVFGFAG-------------------LV 202

Qy        217 GGDAELVRDELAGARIMPDLVFVGLALACPI 247
                |    :|::::|  :  |:  |      |:
Db        203 DGLVGRIREDVSGFSVDHDVAIVATGHTAPL 233


RESULT 12
Q9WZY5
ID   Q9WZY5      PRELIMINARY;      PRT;   246 AA.
AC   Q9WZY5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein TM0883.
GN   TM0883.
OS   Thermotoga maritima.
OC   Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OX   NCBI_TaxID=2336;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MSB8 / DSM 3109;
RX   MEDLINE=99287316; PubMed=10360571;
RA   Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA   Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA   McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA   Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA   Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA   Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT   "Evidence for lateral gene transfer between Archaea and Bacteria from
RT   genome sequence of Thermotoga maritima.";
RL   Nature 399:323-329(1999).
DR   EMBL; AE001754; AAD35964.1; -.
DR   TIGR; TM0883; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   246 AA;  27154 MW;  9E0309AD462CF266 CRC64;

  Query Match           11.6%;  Score 145;  DB 16;  Length 246;
  Best Local Similarity  24.3%;  Pred. No. 0.00037;
  Matches   67;  Conservative   42;  Mismatches  103;  Indels   64;  Gaps   11;

Qy          1 MILELDCGN----------SLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVR 48
              | | :| ||          :  :||        :: |: :::| |   |     :|
Db          1 MYLLVDVGNTHSVFSITEDGKTFRRWR--------LSTGVFQTEDELFSHLHPLLGDAMR 52

Qy         49 ACRLVSVRSEQETSQLVAR--LEQLFPVSALVASSGKQLAGVRNGYLDY-----QRLGLD 101
                : : | |   |   |      :: | :| :   :       :|| : :       :| |
Db         53 EIKGIGVASVVPTQNTVIERFSQKYFHISPIWVKA-------KNGCVKWNVKNPSEVGAD 105

Qy        102 RWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIR 161
              |   :||       |  ::||:||| | |||  :| : || | ||  :|   |   |
Db        106 RVANVVAFVKEYGKNGIIIDMGTATTVDLV-VNGSYEGGAILPGFFMMVHSLFRGT---- 160

Qy        162 YDDAEARRALASLQP-----GQATAE----AVERGCLLMLRGF---VREQYAMACELLGP 209
                  :  |  ::|     |: | |      |  | : |  | |    ::| |
Db        161 -----AKLPLVEVKPADFVVGKDTEENIRLGVVNGSVYALEGIIGRIKEVYG-------- 207

Qy        210 DCEIFLTGGDAELVRDELAGARIMPDLVFVGLALAC 245
              |  : |||| :::|:| :       ||   |:   |
Db        208 DLPVVLTGGQSKIVKDMIKHEIFDEDLTIKGVYHFC 243


RESULT 13
Q8R7M2
ID   Q8R7M2      PRELIMINARY;      PRT;   255 AA.
AC   Q8R7M2;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative transcriptional regulator, homologs of Bvg accessory
DE   factor.
GN   TTE2381.
OS   Thermoanaerobacter tengcongensis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX   NCBI_TaxID=119072;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MB4T / JCM11007;
RX   MEDLINE=21992816; PubMed=11997336;
RA   Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA   Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA   Tan H., Chen R., Wang J., Yu J., Yang H.;
RT   "A complete sequence of T. tengcongensis genome.";
RL   Genome Res. 12:689-700(2002).
DR   EMBL; AE013180; AAM25520.1; -.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   255 AA;  27816 MW;  C3C620ECBC8CA6ED CRC64;

  Query Match           11.5%;  Score 144;  DB 16;  Length 255;
  Best Local Similarity  24.3%;  Pred. No. 0.00047;
  Matches   65;  Conservative   39;  Mismatches  111;  Indels   52;  Gaps    9;

Qy          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDD---------ALVEQLTSQQALPVRACR 51
              |:|  | ||: |   | :|  : :     | |          || ||     : :
```

```
Db          1 MLLAFDVGNTNIVMGVFKG--KKLLHSFRISTDKNKTYDEYGMLVNQLIGYNGISLTEID 58

Qy         52 LVSVRSE----QETSQLVARLEQLFPVSALVASSGKQLAGVRNGYLDYQRLGLDRWLALV 107
                 | : |       | |:::   : |    :|   | :  |:   | : : :| || :  |
Db         59 DVIISSVVPPLMNTLQVMSL--KYFRTKPIVVGPGIK-TGINIKYDNPKEVGADRIVNAV 115

Qy        108 AAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRYDDAEA 167
              ||: |     :||| ||| |   ::  | :||| | ||: :    |   | ::   |
Db        116 AAYELYGGPVIVIDFGTATTFCAISEKGEYLGGIIAPGLMISADALFQRTAKLPKIDLTK 175

Qy        168 ------RRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLG---PDCEIFLTGG 218
                   |  :||:| |             :: | |     :   : |   |   :  |||
Db        176 PPTVINRNTVASMQSG------------IIYGHVGMVDYIVTRMKGEFAPSAYVVATGG 222

Qy        219 DAELVRDE-----------LAGARIM 233
                | :: :|           | | ||:
Db        223 FANMIAEESKTIDTVNEMLTLEGLRII 249


RESULT 14
Q8XHL5
ID   Q8XHL5      PRELIMINARY;      PRT;   259 AA.
AC   Q8XHL5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein CPE2468.
GN   CPE2468.
OS   Clostridium perfringens.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC   Clostridiales; Clostridiaceae; Clostridium.
OX   NCBI_TaxID=1502;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=13 / TYPE A;
RX   PubMed=11792842;
RA   Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA   Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT   "Complete genome sequence of Clostridium perfringens, an anaerobic
RT   flesh-eater.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
DR   EMBL; AP003194; BAB82174.1; -.
DR   InterPro; IPR004619; Baf.
DR   InterPro; IPR000515; BPD_transp.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
DR   PROSITE; PS00402; BPD_TRANSP_INN_MEMBR; UNKNOWN_1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   259 AA;  28819 MW;  0D5FEA3B7A145E10 CRC64;

  Query Match             11.2%;  Score 140;  DB 16;  Length 259;
  Best Local Similarity   23.6%;  Pred. No. 0.001;
  Matches   59;  Conservative   43;  Mismatches  108;  Indels   40;  Gaps    8;

Qy          1 MILELDCGNS-----------LIKWRVIEGAARSVAGGLAESDDALVE--QLTSQQALP 46
                ||| :| ||:           :  ||:   : ::      ||:  ::  || :|  |
Db          1 MILLIDVGNTNIVLGIHDNEKYIASWRISTDSKKT-------SDEYSIQVMQLFNQAKLN 53

Qy         47 VRACRLVSVRSEQETSQLVARLEQL----FPVSALVASSGKQLAGVRNGYLDYQRLGLDR 102
                 : :|      ::  || :    |    :|   | :  |:   | : : :| ||
Db         54 PEDVEGIIISS--VVPNIMHSLENMVRKCFCKEPIVVGPGIK-TGINIKYDNPKEVGADR 110

Qy        103 WLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLGGYICPGMTLMRSQLRTHTRRIRY 162
               :  |||    ||  ::|| ||| |   :   | :||| ||||: :    |      ::
Db        111 IVNAVAAFEKHKKPMIIIDFGTATTFCAITEKGDYLGGNICPGIQISADALFERAAKLPR 170

Qy        163 DDAE------ARRALASLQPGQATAEAVERGCLLMLRGFVREQYAMACELLGPDCEIFLT 216
               : |       :  : |:| |       :  | :  :   |:       :|   :  :  |
Db        171 IELEKPKSVICKNTVTSMQAG------IIYGYIGKVEYIVKRMKKEMMDLGEKEPFVLAT 224

Qy        217 GGDAELVRDE 226
                || |:||  |
Db        225 GGLAKLVYSE 234


RESULT 15
Q9X8N6
ID   Q9X8N6      PRELIMINARY;      PRT;   265 AA.
AC   Q9X8N6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical protein SCO3380.
GN   SCO3380 OR SCE94.31C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Oliver K., Harris D.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL049628; CAB40880.1; -.
DR   InterPro; IPR004619; Baf.
DR   Pfam; PF03309; Bvg_acc_factor; 1.
DR   TIGRFAMs; TIGR00671; baf; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   265 AA;  28203 MW;  E34FA2ABBA946969 CRC64;

  Query Match             11.1%;  Score 138.5;  DB 16;  Length 265;
  Best Local Similarity   23.8%;  Pred. No. 0.0014;
  Matches   70;  Conservative   42;  Mismatches   93;  Indels   89;  Gaps   13;

Qy          1 MILELDCGNSLIKWRVIEGAARSVAGGLAESDDALVEQLTSQQALPVRACRLVSVRSEQE 60
                |:| :| ||:               || : :| :||               :|  | :
Db          1 MLLTIDVGNT-----------HTVLGLFDGED-IVEHWR-----------ISTDSRRT 35

Qy         61 TSQLVARLEQLFPVSALVASSGKQLA---------------------------------- 86
                 :|   |: |  :  |:   | :|
Db         36 ADELAVLLQGLMGMHPLL---GDELGDGIDGIAICATVPSVLHELREVTRRYYGDVPAVL 92

Qy         87 ---GVRNG---YLDY-QRLGLDRWLALVAAHHLAKKACLVIDLGTAVTSDLVAADGVHLG 139
                   ||: |     |: : :| || :  |||  |     :|:| ||| | | |:| | ::|
Db         93 VEPGVKTGVPILTDHPKEVGADRIINAVAAVELYGGPAIVVDFGTATTFDAVSARGEYIG 152

Qy        140 GYICPGMTLMRSQLRTHTRRIRYDDAEARRALASLQPGQATAEAVERGCLLMLRGFVRE- 198
```

```
                | | ||: :    |     ::|  :    |::     |: | ||:: |   :: ||  :
Db        153 GVIAPGIEISVEALGVKGAQLRKIEVARPRSVI----GKNTVEAMQSG---IVYGFAGQV 205

Qy        199 ---QYAMACELLG--PDCEIFLTGGDAELVRDELAGARIM----PDLVFVGLAL 243
                  ||.||     |  :  ||| | :|   |  : ::    | |  :|| |
Db        206 DGVVNRMARELADDPDDVTVIATGGLAPMV---LGESSVIDEHEPWLTLMGLRL 256


Search completed: June 24, 2003, 21:59:35
Job time : 21.2269 secs
```